The Dissertation Committee for Stephanie Ruth Young certifies that this is the approved version of the following dissertation:

Valid Cognitive Ability Measures in the Public Domain: A Convergent Validity Study of the ICAR16 using the WAIS-IV

Committee:

Kevin D. Stark, Supervisor

Timothy Z. Keith, Co-Supervisor

Tiffany Whittaker

Adreana Haley

Valid Cognitive Ability Measures in the Public Domain: A Convergent Validity Study of the ICAR16 using the WAIS-IV

by

Stephanie Ruth Young

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2020

Dedication

For my dad, who told me to do my best and try my hardest.

Acknowledgements

Thank you to my supportive committee members for making this project possible. I cannot thank my advisor and mentor, Dr. Keith, enough for his calm and thoughtful guidance throughout my time in graduate school. I'd also like to thank Dr. Stark, who has been a constant source of encouragement in my clinical and academic work. Thank you so much to my labmates and cohort members who contributed their time to administer the WAIS-IV to what felt like a lot of participants, despite what the limitations section says. Finally, thank you so much to my loving family for supporting me on this long journey, and to Rathan, for listening to me read these pages so many times. I could not have done this without you.

Abstract

Valid Cognitive Ability Measures in the Public Domain: A Convergent Validity Study of the ICAR16 using the WAIS-IV

Stephanie Ruth Young, Ph.D.

The University of Texas at Austin, 2020

Supervisor: Kevin D. Stark

Co-supervisor: Timothy Z. Keith

One of the main barriers to the utility of the International Cognitive Ability Resource (ICAR) is that no research to date has been conducted on its construct validity using a theory – based individual cognitive assessment battery. The aim of the current study was to examine the relations between the ICAR16 items and the Cattell-Horn-Carroll (CHC) broad abilities as measured by the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV). Research questions include: 1) Is there evidence of the convergent validity of the ICAR16 when compared with a gold-standard individual assessment of cognitive abilities? And, 2) Which CHC constructs are related to the respective ICAR subtests?

Data was collected from a convenience sample of university student volunteers (N = 67) and a clinical sample from a university-based assessment center (N = 30) who completed the ICAR Sample Test (ICAR16) and the ten core subtests of the WAIS-IV. To address the first research question, the correlations between the confirmatory-factor estimated general factors from the ICAR16 and WAIS-IV were examined. The model fit the data well ($\chi 2$ (19) = 14.15, p= .78) and revealed a large correlation between the general factors (r =.94, p<.001). The range-and-reliability-corrected correlation between the WAIS-IV FSIQ and the ICAR16 total score was also large ($r_{ICAR16,\ FSIQ}$= .81 p<.001).

To address the second research question, correlational methods of examining convergent and discriminant validity were employed. Evidence from range-and-reliability-corrected

correlations suggests that the ICAR16 Letter-Number Series task is most closely related to fluid reasoning (Gf) ($r=.70$, $p<.01$) while the Matrix Reasoning, Verbal Reasoning, and Three-Dimensional Reasoning tasks are most closely related to visual-spatial reasoning (Gv) ($r=.35$-$.75$). As a point of discriminant validity, all of the ICAR tasks demonstrated the lowest correlations with measures of processing speed (Gs) or working memory (Gsm) ($r=.07$-$.32$).

Findings suggest that the ICAR16 provides a valid estimate of nonverbal intelligence. Results of the present study suggest that the ICAR16 may not be sensitive enough to discriminate between distinct CHC abilities, though more specific associations may be revealed in a larger sample. The present study provides a foundation for future validation and use of the ICAR.

Table of Contents

Table of Tables

Table of Figures

Chapter 1: An Examination of Convergent Validity of the ICAR16 using the WAIS-IV

1. Introduction

Cognitive abilities are among the strongest known predictors of academic achievement (Deary, Strand, Smith, & Fernandes, 2007; McGrew & Wendling, 2010; Rohde & Thompson, 2007), occupational success (Kuncel & Hezlett, 2001; Schmidt & Hunter, 1998), and individual health and well-being (Batty & Deary, 2004; Batty, Deary, & Gottfredson, 2007; Deary, Batty, & Gale, 2008), which explains the commercial success and widespread use of proprietary tests of cognitive abilities such as the Woodcock-Johnson, Wechsler, and Kaufman batteries. Commonly referred to as intelligence quotient (IQ) tests, proprietary cognitive ability assessments have historically targeted educational, clinical, and industrial settings (Elliott, 2007; Kaufman, 2004; Schrank, Mather, & McGrew, 2014; Wechsler, 2014). The primary focus of these assessments is to identify patterns of strengths and weaknesses within individual examinees to provide meaningful feedback and guide decision making-processes in practical contexts.

Although commercial assessments are mainly designed for and marketed towards end-users (e.g., school psychologists, neuropsychologists, human resources personnel), primary researchers tend to default to these tools as well. For example, in a research synthesis of 134 analyses of intelligence data, 126 analyses used Woodcock-Johnson cognitive assessment products, and the remaining studies used other commercial measures (McGrew & Wendling, 2010). Continuity of measurement between research and practice offers obvious benefits in applied clinical fields, yet the measurement needs of primary researchers often diverge from those of commercial test users, particularly for those in nonclinical fields. Researchers interested in evaluating the relation between cognitive abilities and a range of other constructs tend to be less concerned with interpretive feedback than clinical test users, for example (Condon & Revelle, 2014).

Unfortunately, most existing proprietary assessments are expensive to access, require trained administrators, and are prohibitively time-consuming to administer in the context of a research study (see Camara, Nathan, & Puente, 2000). Limited accessibility to measurement tools may prevent the inclusion of important cognitive ability variables in studies across a variety of fields, and in turn stifle research progress (Gambardella & Hall, 2006; Goldberg, 1999). Flexible, well-validated tools in the public domain can help make cognitive ability measurement more accessible for researchers across fields.

1.1 International Cognitive Ability Resource (ICAR)

To address the lack of accessibility of cognitive assessment in primary research, Condon and Revelle (2014) developed the International Cognitive Ability Resource (ICAR), a public-domain assessment tool with four published item types, and several other types under development. The aim of the ICAR project is to encourage broader assessment of cognitive abilities in the social sciences and healthcare fields by providing flexible, unrestricted test items to researchers (Revelle et al., 2014). Although a wealth of well-established commercial cognitive ability measures exists, free measures tend to be proprietary or not well validated. For example, the most common freely available battery, the ETS Kit, lacks sufficient evidence of its construct validity, and its usefulness in primary research has been called into question (Babcock & Laguna, 1997). Higher quality assessment tools are beginning to emerge to address the accessibility problem of cognitive measurement in research. Notably, the NIH toolbox offers a validated cognitive battery at a low cost to researchers (Akshoomoff et al., 2014), however, this tool is not freely available in the public domain. The ICAR is the first non-proprietary resource that freely distributes items to qualified researchers and encourages researchers to develop and contribute items for use and validation by others.

Initial analyses provide evidence of the reliability, factor structure, and predictive validity of the four item types from a large, international sample of individuals ages 14 to 90 who completed the assessment online (Condon & Revelle, 2014; Young, Keith, & Bond, 2020). The convergent validity of the ICAR items was also evaluated against the *Shipley-2*, another brief, self-administered measure of cognitive abilities (Shipley et al., 2009). Shared characteristics (e.g. self-administration, brevity) make the measure a useful point of comparison for the ICAR, however there has been little research on the construct validity of the *Shipley-2* (Reynolds et al., 2016) and the few studies that have evaluated the measure have raised concerns about its usefulness (Beaujean et al., 2017; Reynolds et al., 2016). One of the main barriers to the adoption of the ICAR is that no research to date has been conducted on its construct validity using a gold-standard, individually-administered cognitive battery.

1.2. Cattell-Horn-Carroll Theory

Many contemporary tests of intelligence have either been developed upon or have been adapted to fit within the Cattell-Horn-Carroll (CHC) theory of cognitive abilities (Keith & Reynolds, 2010). Although alternative theoretical frameworks for intelligence exist (see Das, Naglieri, & Kirby, 1994; Gardner, 1987; Johnson & Bouchard, 2005), CHC theory offers the most empirical support, is most frequently applied in assessment, and is considered to be the common language to communicate intelligence test scores and research findings (Flanagan & Harrison, 2012; Keith & Reynolds, 2010; McGrew, 2005, 2009).

CHC theory proposes a three-stratum hierarchical model of intelligence, with one superordinate general intelligence factor at the apex that influences several broad abilities, which in turn influence even more narrow abilities (Carroll, 1993). Each of these abilities have important implications, but a few broad abilities tend to stand out in terms of their factor cohesion, and strength of their association with general intelligence (Table 1) (McGrew & Wendling, 2010; Taub, Keith, Floyd, & Mcgrew, 2008). See the Integrated Analysis chapter for an in-depth discussion of CHC and competing theories of intelligence.

Table 1. Selected CHC Broad Abilities Commonly Measured by Cognitive Ability Tests

CHC Broad Ability	**Description**
Fluid Reasoning (Gf)	The ability to use deliberate and controlled mental operations to solve novel problems. May be viewed as the aptitude to learn and drives the development of other abilities. Strong relation with general intelligence.
Verbal Comprehension (Gc)	The knowledge of cultural information acquired primarily through the process of acculturation. Primarily language based.
Visual-Spatial Ability (Gv)	The ability to generate, store, retrieve, and transform visual images and tactile sensations.
Working Memory (Gsm)	The ability to apprehend and maintain awareness of a limited number of elements of information in the immediate situation. Comprised of memory span and working memory. Based on the Baddley model of working memory (Baddeley, 2001)
Processing Speed (Gs)	The ability to automatically and fluently perform simple or over-learned cognitive tasks, especially when a high degree of focused attention is required.

Note. Descriptive note. Adapted from McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1-1.

Among the broad abilities, fluid intelligence (i.e. the ability to solve novel problems; Gf) and crystallized intelligence (i.e. stored knowledge and its retrieval; Gc) tend to correlate most highly with general intelligence (Schneider & McGrew, 2012; Schrank, Decker, & Garruto, 2016). Gf, in particular, is often viewed as an approximation of general intelligence (Aken, Kessels, Wingbermühle, Veld, & Egger, 2016; Gustafsson, 1988; Kvist & Gustafsson, 2008; Schneider & Newman, 2015) and its measurement tends to be relatively less culturally-dependent than other broad abilities (Arvey, 1972; Raven, 2000; Schneider & Newman, 2015). Gf has also been shown to have a strong relation to working memory (Gsm) (Au et al., 2015;

Martínez & Colom, 2009), and there is some evidence Gf is the driver of the accumulation of acquired knowledge (Gc) and visual-spatial ability (Gv) (Bigler, Johnson, Jackson, & Blatter, 1995; Gong et al., 2005; Martínez & Colom, 2009). For these and other reasons, Gf has been the focus of many brief measures of intelligence (Hossiep, Turck, & Hasella, 1999; Johnsen, 2017; Kvist & Gustafsson, 2008; J. Raven, 2000). The content of the ICAR items suggests Gf is also the main construct measured, but research has yet to confirm this idea.

The Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV), is one of the most commonly used and well-validated measures of intelligence and of CHC abilities, and appears to measure four to five CHC broad abilities well (Benson, Hulac, & Kranzler, 2010; Niileksela, Reynolds, & Kaufman, 2013; Weiss, Keith, Zhu, & Chen, 2013). Thus, an investigation of how the ICAR items relate to the WAIS-IV provides evidence of construct validity for the existing ICAR item set, as well as build a foundation for the development of additional items and future construct validity research.

Table 2. WAIS-IV Subtests, Associated CHC Abilities, Average Reliability, and Broad Ability Factor Loadings

Subtest (Acronym)	**Task Description**	***CHC Ability***	***Average r_{xx}***	***Factor loading***
Arithmetic (AR)	Solve math word problems presented verbally.	Gf	.88	.72
Block Design (BD)	Construct a visually displayed pattern with three-dimensionally colored blocks.	Gv	.87	.77
Coding (CD)	Translate strings of numbers to figures designated in a key under a time constraint.	Gs	.86	.90
Digit Span-Forward (DS-F)	Recite a series of numbers presented orally verbatim	Gsm	.81	.56
Digit Span-Backward (DS-B)	Recite a series of numbers presented orally in reverse order	Gsm	.82	.72
Digit Span – Sequential (DS-S)	Recite a series of numbers presented orally in sequential order	Gsm	.83	.73
Information (I)	Answer general information questions presented verbally about a variety of topics.	Gc	.93	.81
Matrix Reasoning (MR-W)	Select the image that best completes the pattern presented visually.	Gf	.90	.74
Similarities (SI)	Describe how two words presented orally are similar.	Gc	.87	.86
Symbol Search (SS)	Mark the presence or absence of a target figure in a series of figures under a time constraint.	Gs	.81	.78
Visual Puzzles (VP)	Select three shapes to create a larger two-dimensional shape.	Gv	.89	.70
Vocabulary (VC)	Define words presented orally.	Gc	.94	.87

Source: Essentials of WAIS-IV Assessment, Second Edition by E. O. Lichtenberger, and A. S. Kaufman, 2012, p. 34-38. Copyright 2012 by John Wiley & Sons Inc.

The aim of the current study was to examine the relations between the ICAR items and the WAIS-IV. Scores on the assessments from a convenience sample of university student volunteers and a clinical sample from a university-based psychological assessment center were analyzed and interpreted. Bivariate correlations and factor analytic methods were used to examine the following research questions: 1) How does the ICAR16 compare to the WAIS-IV as an overall estimate of general intelligence? And, 2) How do the respective ICAR item types relate to the CHC broad abilities as measured by the WAIS-IV?

2. Method and Materials

2.1. Sample

Data were collected from two samples: an existing database of 30 university students ages 18 to 28 who sought assessment services at a university-based assessment center, and a convenience sample of 67 student volunteers ages 18 to 45 from the same large, southwestern university. Volunteers were recruited through the university's college of education subject pool and word of mouth. *A priori* participation criteria excluded participants who completed a WAIS-IV or an ICAR16 within in the last year, and participants with WAIS-IV Full Scale Intelligence Quotient (FSIQ) scores of less than 70, though no such cases presented in either sample.

The clinical sample differed slightly from the volunteer sample in regard to age, FSIQ on the WAIS-IV, ICAR16 total score, and year in university (Table 3). Participant age, FSIQ scores, and total scores on the ICAR16 were lower in the clinical sample, though within one population-based standard deviation of the volunteer sample (WAIS-IV norming sample standardized mean=100, S.D.= 15; ICAR16 norming sample mean=8.05, S.D.= 3.73). Furthermore, the average FSIQ score obtained by participants in the clinical sample was slightly higher than those of the respective norming samples (Condon & Revelle, 2014; David Wechsler, 2008). Thus, differences in the two samples were statistically significant but not clinically meaningful, and the combination of the two samples does not substantially threaten the validity of analyses for the purposes of this study.

Table 3. Volunteer, Clinical, and Combined Sample Demographics

	Combined Sample	Volunteer Sample	Clinical Sample	t-test statistic	*p*-value
N (Female)	97 (51)	67 (32)	30 (19)		
Age	22.47 (4.08)	23.06 (4.37)	21.17 (3.02)	2.46	0.02
FSIQ	112.94 (11.20)	116.09 (10.71)	105.59 (8.93)	4.87	<.01
ICAR16	9.82 (3.83)	10.45 (3.85)	8.43 (3.47)	2.55	0.01

Note: Age, FSIQ, and ICAR16 statistics presented as Mean (Standard Deviation).

While the combined sample was roughly evenly distributed by gender, participants were largely homogenous in age, racial/ethnic identity, and education level (Table 4). The majority of participants who provided race and ethnicity information identified as white and non-Hispanic. Due to the sampling process, all students had at least some higher education experience. WAIS-IV and ICAR16 norming samples were also made up of majority white and non-Hispanic participants, but offer more diversity in regard to age and education level.

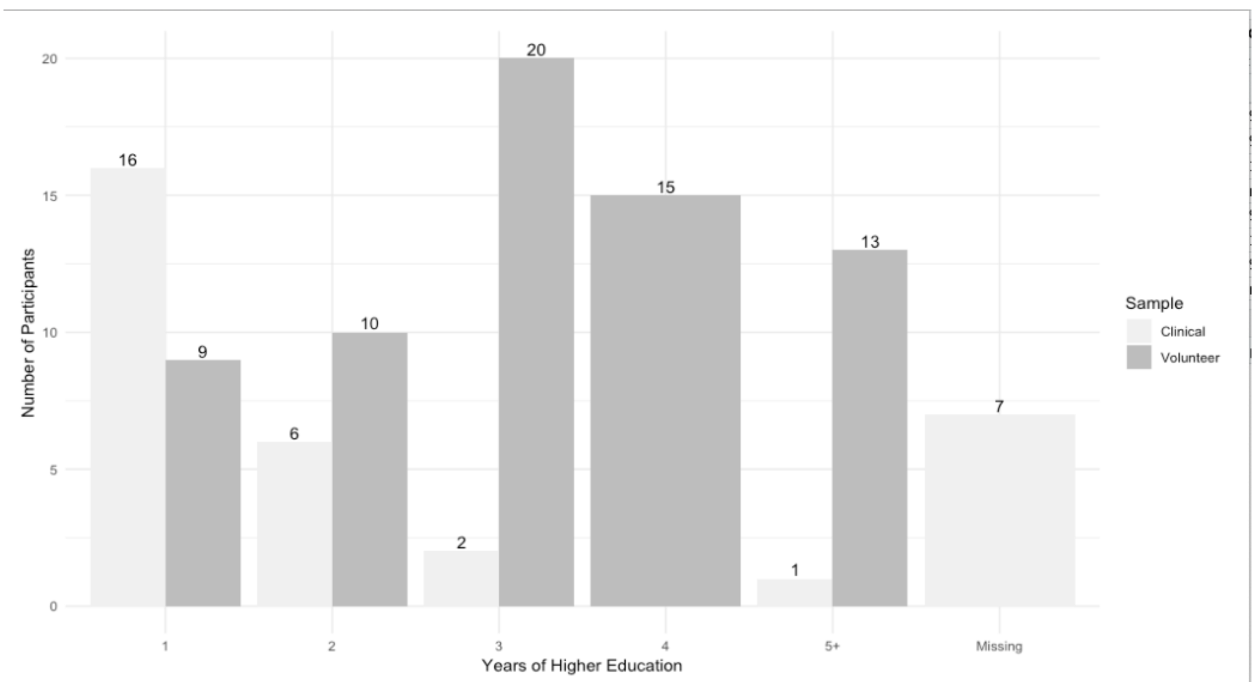


Figure 1. Years in University by Sample Type

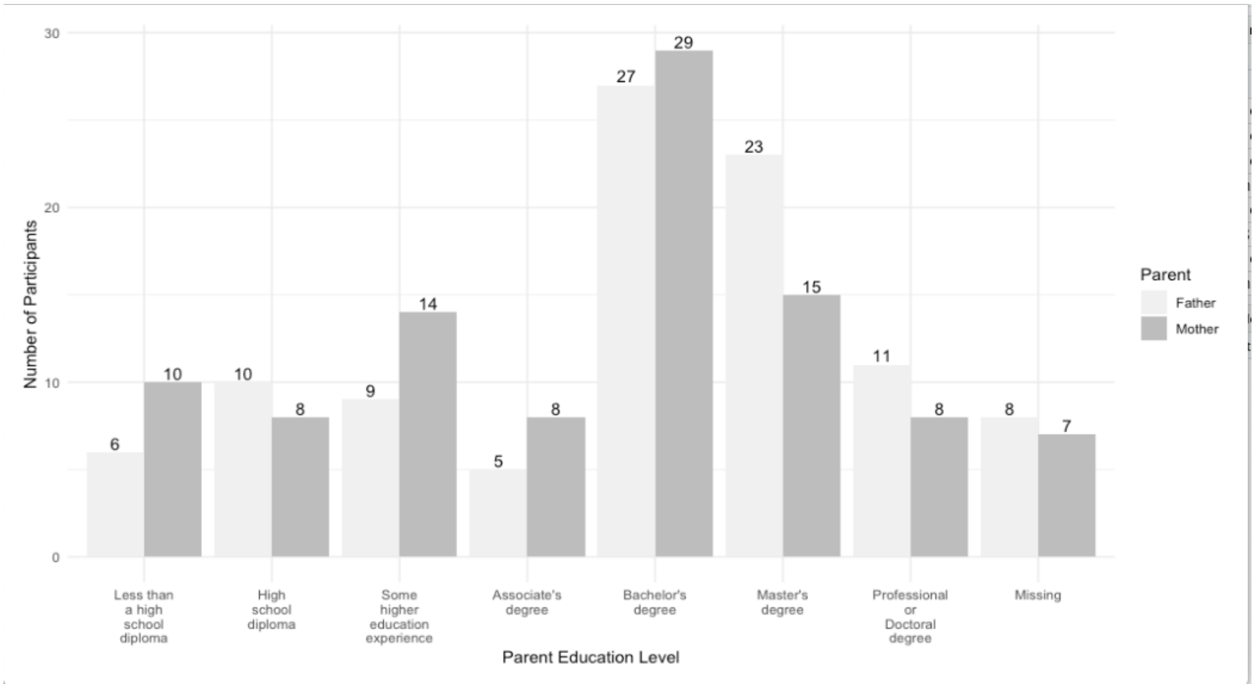


Figure 2. Parental Education Level

Table 4. Comparison of Demographic Information of Respective Norming Samples

	Present Study (N=97)	WAIS-IV (N= 2200)	ICAR* (N=75740)
Age			
Mean (SD)	22.47 (4.08)	42.26 (13.82)	25.71 (10.64)
Range	[18.32, 45.57]	[16, 90]	[14, 90]
Female %	51	50	68
Racial Identity %			
White	59	70	68
Asian	29	4	10
Black or African American	3	12	8
Other	6	14	14
No information	--	--	--
Ethnic Identity %			
Hispanic (Any race)	16	14.5	7.8
Not Hispanic	50	85.5	
No information	31	--	--
Education %			
<12 years	0	16	16
12 years	34	31	6
13-16 years	58	28	69
>16 years	5	25	9
No information	--	--	--

Note. *Race/Ethnicity information is not included in the ICAR norming data set and instead was gathered from Condon and Revelle (2012). Race/ethnicity Information is based on U.S. participants only at the time of publication and may not be reflective of demographics of the full sample. Demographic information about the WAIS-IV sample was estimated based on statistics in the *WAIS-IV Technical and Interpretive Manual* (Wechsler, 2008)

2.2 Measures

2.2.1 Demographics Questionnaire

The demographics questionnaire was a brief online survey administered through Qualtrics. The survey instrument collected information about age, gender, race, ethnicity, parental education, and college education. The demographic survey is included in Appendix D.

2.2.2 Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV).

The WAIS-IV is an individually administered test of intelligence for adults ages 16 to 90 (Wechsler, 2008a). It contains 10 core subtests (see Table 2) which yield four composite indices, a Full-Scale IQ (FSIQ) score, and a General Ability Index (GAI). The Full-Scale IQ (FSIQ) score is derived by summing scaled scores reflecting performance with 10 core subtests and corresponds to psychometric *g*. The GAI is an alternative estimate of general intelligence that excludes measures of Gsm and Gs (i.e. AR, DS, CD, and SS). The WAIS-IV also includes four composite scores that reflect broad ability areas: the Verbal Comprehension Index (VCI, a measure of Gc); the Perceptual Reasoning Index (PRI, a measure of Gf and Gv); the Working Memory Index (WMI, a measure of Gsm); and the Processing Speed Index (PSI, a measure of Gs) (Weiss et al., 2013).

The WAIS-IV Technical and Interpretive Manual reports the normative sample included 2,200 individuals and was demographically representative of the U.S. population based on age, gender, ethnicity, geographic region, and education. Internal consistency was moderate to high (r = .71 to .96) for the individual subtests, and very high for the Full-Scale IQ (r = .97 to .98) and the GAI (r=.96 to .98) (Table 2). Moderate to high correlations between the WAIS-IV subtests and the Wechsler Individual Achievement Test, 2nd edition (r = .42 to .80), and with the previous edition of the WAIS (r = .33 to .81) provide evidence of predictive and concurrent validity, respectively (Wechsler, 2008).

2.2.3 International Cognitive Ability Resource-16 item (ICAR16).

The 16-item subset of the 60-item International Cognitive Ability Resource was published by Condon and Revelle (2014) as a sample test and is included in Appendix E. The ICAR16 items are self-administered, multiple choice, and untimed. The respective multiple choice item types ask participants to identify: the missing element in a pattern of digit or letter combinations (Letter Number Series, LN, 9 items), the missing shape in a 3x3 array of geometric shapes (Matrix Reasoning, MR-I, 11 items), the correct response to a variety of logic questions and math word problems (Verbal Reasoning, VR, 16 items), and the response choice that is a possible rotation of a stimulus figure (Three-dimensional Rotation 3DR, 24 items) (Condon & Revelle, 2014).

The ICAR16 includes four items from each of the four item types. Table 2 provides the reliability coefficients and the estimated factor loading on the general factor for each item type from a subset of the online sample consisting of 1909 individuals. This subset was 72% female with a mean age of 19.7 (SD=1.2). Information about nationality, race, ethnicity, and education were not provided about this subsample. Internal consistency was acceptable for each of the four types except for Matrix Reasoning (MR-I), which was only marginally acceptable. The average reliability of ICAR16 was adequate (α= .81; ω total=.83) while the full 60-item set was good (α= .93; ω total= .94). As expected, the long-form version demonstrated higher reliability, however, slightly more variance was extracted from the ICAR16 by the general factor (ICAR16 ω hierarchical= .66; ICAR60 ω hierarchical = .61).

Table 5. ICAR16 Online Sample Internal Consistency Estimates and General Factor Loadings

	Coefficient Alpha α	Omega Hierarchical ωh	Omega Total ωt	General Factor Loading
LN	.68	.63	.71	.8
MX	.52	.67	.56	.7
3DR	.74	.49	.76	.5
VR	.59	.72.	.62	.7
ICAR16	.81	.57	.83	

The ICAR16 was selected for use rather than the ICAR60 for several reasons. The first aim was to replicate the concurrent validity assessment procedures from Condon and Revelle (2014) Study 3 in order to interpret results across studies. Additionally, the sampling style of Condon and Revelle (2014) was such that participants completed random subsets of the ICAR60, while the ICAR16 sample test was administered in its entirety to a subset of participants. Another advantage to validating the ICAR16 is that although all 60 items are available upon authorization, researchers must apply to use the items, and they are not yet packaged together. Alternatively, the ICAR16 sample test was published in *Intelligence* as a supplementary material to Condon and Revelle's (2014) article introducing the measure. The ICAR16 also offers sound psychometric properties but in a briefer, more accessible format than the ICAR60, and is likely to appeal to a broad group of researchers.

2.3 Procedure

In the clinical sample, participants who sought assessment services through a university-based assessment clinic completed an intake survey via email to collect relevant demographic and educational information. Within this survey, participants were provided a detailed description of the study, which explained the study purpose and indicated their consent to have their de-identified data used for research purposes upon completion of testing (Appendix C). Consenting partipants then completed the 16-item ICAR Sample Test within the same Qualtrics survey. Participants were administered the ten core subtests of the WAIS-IV within a larger battery of neuropsychological assessments.

Students who volunteered to participate apart from the clinical sample were administered the same consent forms, demographic survey, the ICAR16 items, and the ten core subtests of the WAIS-IV within the same session. All participants were assessed on the WAIS-IV by trained graduate student examiners in the university assessment center and supervised by a licensed psychologist. After data collection, all data were stripped of identifying information and combined for analysis.

2.4 Analysis

Correlational analyses were conducted in RStudio (RStudio Team, 2019) with the psych package (Revelle, 2019). To address the first research question (i.e., How does the ICAR16 relate to the WAIS-IV as an overall estimate of general intelligence?) the correlation between the CFA-estimated general factors from the ICAR16 and WAIS-IV was examined and compared to those between other brief, homogenous assessments of cognitive ability tests with traditional intelligence batteries. The general factor for the ICAR was indicated by the total scores from the four subtests. The general factor on the WAIS-IV was indicated by the four main composite indices (VCI, PRI, WMI, and PSI). Pearson correlation coefficients between the total score on the ICAR16 and the FSIQ and GAI observed scores were also examined.

To address the second research question (i.e. which CHC constructs are related to which ICAR item types?), correlations between variables hypothesized to measure the same construct were compared with those thought to measure disparate constructs (Campbell & Fiske, 1959). An *a priori* power analysis was conducted to determine the minimum sample size necessary to obtain Cohen’s (1988) recommended power of .8. A moderate effect size of $r=.3$ was selected to reflect the smallest correlations typically observed in concurrent validity studies using the WAIS-IV and batteries similar in length and nature to the ICAR (Balboni, Naglieri, & Cubelli, 2010; Bell, Rucker, Finch, & Alexander, 2002; Krach, Loe, Jones, & Farrally, 2009; Lodge, 2012; Salthouse, 2009). A two-tailed bivariate normal distribution power analysis with an alpha error probability of .05 revealed a sample size of 84 participants is necessary to detect the effect size desired, which is achieved with the current sample.

To represent the CHC broad abilities, broad ability composite (BAC) scores were created using the interpretive system presented by Lichtenberger and Kaufman (2012) (Table 6). The following Pearson correlation coefficients were calculated: 1) the correlation between the respective BAC scores and the ICAR item type composite scores, 2) the correlations between the individual WAIS-IV subtest scores and the ICAR item type composite scores, 3) the correlations between the FSIQ score and the ICAR16 total scores, 4) the correlation between the GAI score and the ICAR16 total score.

Table 6. Lichtenberger and Kaufman (2012) System for Interpreting CHC Broad Abilities

Factor Name	**Core Five-Factor Model**	**Internal Consistency Coefficient (α)**
Crystallized Intelligence (Gc)	Vocabulary + Information	.96
Short-Term Memory (Gsm)	Digit Span Backwards + Digit Span Sequencing	.88
Fluid Reasoning (Gf)	Matrix Reasoning + Arithmetic	.93
Visual Processing (Gv)	Block Design + Visual Puzzles	.93
Processing Speed (Gs)	Symbol Search + Coding	.90

Note. Descriptive note. Adapted from *Essentials of WAIS-IV Assessment, Second Edition* by E. O. Lichtenberger, and A. S. Kaufman, 2012, p. 166-169. © 2012 by John Wiley & Sons Inc.

Because the sample is homogeneous in terms of education attainment (i.e. all participants had been admitted to a selective university), correction methods for restriction of range and reliability were replicated from Condon and Revelle (2014) Study 3. Bryant and Gohale (1972) and Alexander (1990) provide formula (1) for correcting restricted correlations between two variables, x and y, selected on a third unmeasured variable:

$$\hat{r}_{xy} = \text{rxy(sx/Sx)(sy/Sy)} \pm \sqrt{(1-(s_x/S_x)^2)(1-(s_y/S_y)^2)} \quad (1)$$

where $\hat{r}_{xy}$ is the correlation between the item scores on the respective tests corrected for restriction of range, s_x and s_y are the standard deviations in the restricted sample (i.e. the study sample), and S_x and S_y are the standard deviations in the unrestricted sample (i.e. the norming sample).

Due to the relatively small sample size of the study, published reliabilities (Wechsler, 2014) for the WAIS-IV subtests as well as the ICAR16 item type scores (Condon & Revelle, 2014) were used instead of the sample reliabilities. Formula (2) was applied as a correction to provide a more accurate estimation of the correlations by accounting for measurement error (Murphy & Davidshofer, 1988):

$$\hat{r}_{x'y'} = \frac{r_{xy}}{\sqrt{(r_{xx})(r_{yy})}} \qquad (2)$$

where $\hat{r}_{x'y'}$ is the correlation between the WAIS-IV subtest scores and the ICAR16 composite scores corrected for reliability, rxy is the correlation of the variables in the present study sample, and rxx and ryy are the reliabilities of the variables from their respective norming samples.

Correlations between ICAR item type composites and WAIS-IV subtest scores were compared to evaluate if there is a statistically significant difference between ICAR16 subtest scores and the (hypothesized) related and unrelated WAIS-IV subtests, respectively. For this, a z-test for inequality of two dependent Pearson r's, also known as a Steigler's z-test for correlated correlations within a population (Steiger, 1980), was employed. The formula is as follows:

$$z = zwx - zwy * \frac{\sqrt{N-3}}{\sqrt{2*(1-rxy)*h}} \qquad (3)$$

where $h = \frac{1-(f*rm^2)}{1-rm^2}$ (4) and $f = \frac{1-rxy}{2*(1-rm^2)}$ (5) and $rm^2 = \frac{rwx^2+rwy^2}{2}$ (6)

and where r_{wx} and r_{wy} represent the correlations between a given ICAR composite and a hypothesized related WAIS-IV subtest and the same ICAR composite with a hypothesized unrelated WAIS-IV subtest. r_{xy} represents the common index correlation, the correlation between the unrelated WAIS-IV subtest and the related WAIS-IV subtest in question. Subtest and composite index correlations were sourced from the *WAIS-IV Technical and Interpretive Manual*. z_{wx} and z_{wy} represent the Fischer's z transformations of r_{wx} and r_{wy} respectively.

As a point of reference for discrimination, each correlation between the ICAR composite scores and the related BAC scores were compared to correlations between those of the latter and the Gs BAC score. Gs was chosen as a point of discrimination for two reasons. First, Gs is not explicitly measured on the ICAR as the items are untimed. Second, measures of Gs demonstrate the lowest correlations with the WAIS-IV subtests that measure the constructs of interest on the ICAR, namely Gf and Gv (Wechsler, 2008).

3. Results

3.1. Research Question 1: *Is there evidence of convergent validity of the ICAR16 when compared with a well-established, long-form measure of intelligence?*

Figure 2.

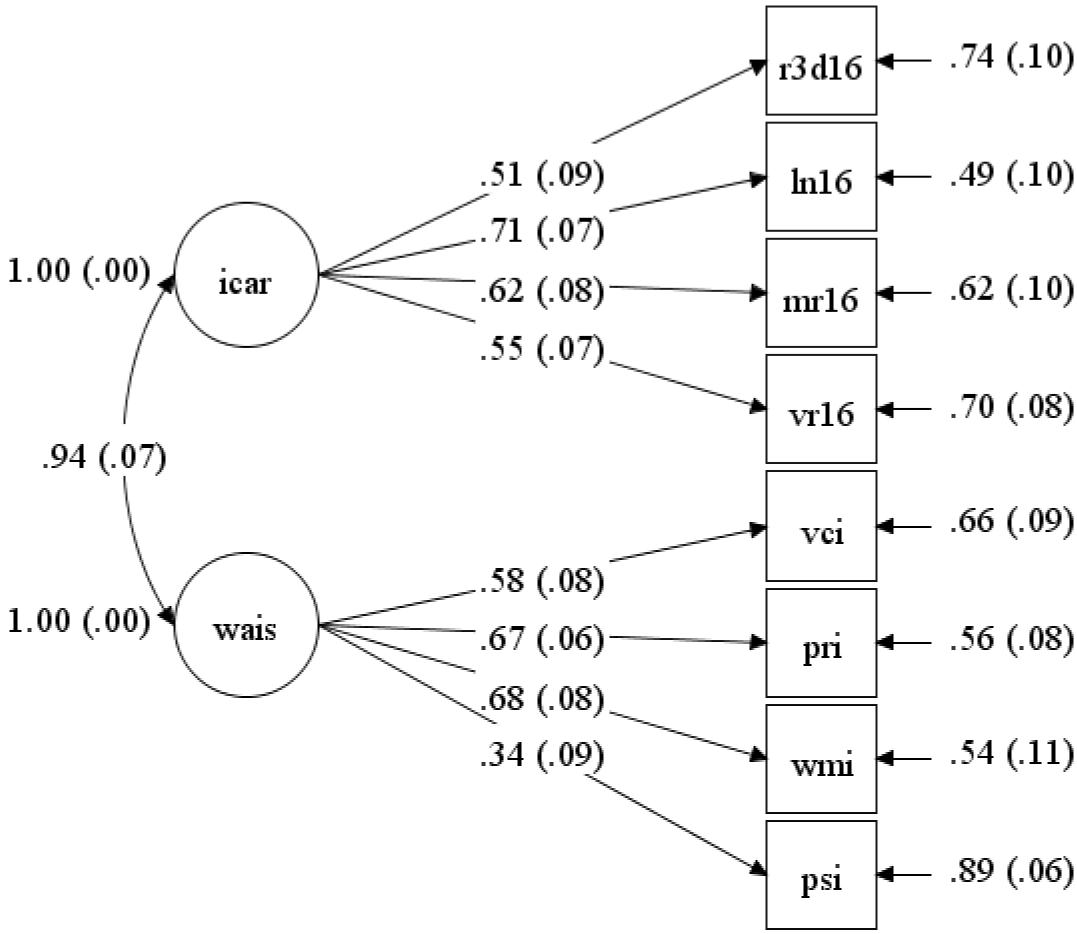


Figure 3. Correlated Two-Factor Model

Correlations between ICAR general factor and WAIS general factor were expected to be large, similar to other tests of abilities (r > .7). The two-factor model (Figure 2) fit the data well ($\chi 2$ (19) = 14.23, *p*= .77; RMSEA < .001, 90% C.I. (.000, .062); CFI = 1.000; TLI = 1.045) and revealed a large correlation between the general factors (*r* =.94, *p*<.001).

Correlations between the observed general intelligence scores were also hypothesized to be large (r>.7), though smaller than those between the general factors. Pearson correlations were calculated for the observed general scores on the WAIS-IV, the FSIQ and the GAI, and the ICAR16 total score respectively. Uncorrected correlations were significant but did not reach the hypothesized magnitude of r=.7. Range and reliability corrected correlations were slightly higher than uncorrected correlations and exceeded the expected magnitude. Contrary to expectations, neither uncorrected nor corrected correlations demonstrated any statistically significant differences in correlations between the ICAR16 total scores and the FSIQ and GAI, respectively.

Table 7. ICAR16 Total Score Correlations with WAIS-IV General Ability Composite Scores

	Pearson Correlation [95% CI]		Steigler's Difference Test	
ICAR16 Total Score	FSIQ	GAI	z	p-value
Uncorrected Correlations	.62** [.49, .73]	.61** [.49, .73]	.26	.79
Corrected for Restriction of Range	.72** [.62, .81]	.71** [.61, .80]	.41	.69
Corrected for Restriction of Range and Reliability	.81** [.74, .88]	.81** [.74, .87]	.21	.84

Note. * $p < .05$, ** $p < .01$. Restriction of range was corrected by using formula (1) from Bryant and Gohale (1972) and Alexander (1990). Formula (2) from Murphy and Davidshofer (1988) was then applied to the range-corrected correlations to using the reliabilities published in the WAIS-IV technical manual (David Wechsler, 2008) and the initial ICAR validation study (Condon & Revelle, 2014).

3.2. Research Question 2: *Which CHC Constructs are related to the ICAR types?*

Correlations between the raw broad ability composite (BAC) scores and ICAR16 subtests and confidence intervals were calculated, and then sequentially corrected for range and then reliability to replicate the correction process of Condon and Revelle (2014) (Table 7, 8, 9). Uncorrected, range corrected, and range-and-reliability corrected correlations between ICAR16 subtests and the ten core subtests on the WAIS-IV are provided in Appendix B. As expected, visual inspection of the correlations revealed each ICAR16 subtest correlated the highest with Gf or Gv, and the lowest with Gsm or Gs. Steigler's z tests were conducted to determine if the magnitudes of the correlations with the BAC scores were statistically different from one another for each respective ICAR subtest.

Table 8. Uncorrected Correlations Between Raw Broad Ability Composites and ICAR16 Subtests

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Gf	37.65	5.38								
2. Gv	66.27	13.38	.39** [.21, .55]							
3. Gc	61.83	9.53	.38** [.19, .54]	.18 [-.02, .37]						
4. Gsm	18.68	4.01	.41** [.23, .57]	.17 [-.03, .36]	.35** [.16, .52]					
5. Gs	115.08	20.75	.27** [.07, .44]	.25* [.05, .42]	.14 [-.06, .34]	.29** [.09, .47]				
6. LN16	2.42	1.54	.56** [.40, .68]	.35** [.16, .51]	.34** [.15, .51]	.33** [.14, .50]	.32** [.13, .49]			
7. MR16	2.51	1.31	.37** [.18, .53]	.39** [.21, .55]	.24* [.04, .42]	.30** [.11, .48]	.22* [.02, .40]	.42** [.24, .57]		
8. VR16	3.20	0.95	.27** [.08, .45]	.46** [.29, .60]	.22* [.02, .41]	.30** [.10, .47]	.08 [-.13, .27]	.39** [.21, .55]	.34** [.15, .50]	
9. R3D16	1.70	1.47	.20 [-.00, .38]	.33** [.14, .49]	.31** [.11, .48]	.19 [-.02, .37]	.06 [-.14, .26]	.41** [.23, .56]	.39** [.21, .55]	.19 [-.01, .38]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

Table 9. Range Corrected Correlations Between BAC scores and ICAR16 Subtests

Variable	Gf	Gv	Gc	Gsm	Gs	LN16	MX16	VR16	R3D16
1. Gf	1								
2. Gv	.39**	1							
3. Gc	.51**	.26*	1						
4. Gsm	.41**	.17	.35**	1					
5. Gs	.26**	.26**	.14	.29**	1				
6. LN16	.56**	.35**	.34**	.33**	.34**	1			
7. MX16	.37**	.39**	.24*	.30**	.23*	.41**	1		
8. VR16	.34**	.55**	.28**	.37**	.10	.46**	.42**	1	
9. R3D16	.17	.29**	.27**	.16	.07	.38**	.35**	.16	1

Note. Restriction of range was corrected by using formula (1) from Bryant and Gohale (1972) and Alexander (1990). manual (David Wechsler, 2008) and the initial ICAR validation study (Condon & Revelle, 2014). Values in square brackets indicate the 95% confidence interval for each range corrected correlation. * indicates $p < .05$. ** indicates $p < .01$.

Table 10. Range and Reliability Corrected Correlations Between BAC Scores and ICAR16 Subtests

Variable	Gf	Gv	Gc	Gsm	Gs	LN16	MX16	VR16	R3D16
1. Gf	.93								
2. Gv	.42**	.93							
3. Gc	.54**	.27**	.96						
4. Gsm	.46**	.19	.38**	.88					
5. Gs	.29**	.27**	.16	.33**	.90				
6. LN16	.70**	.44**	.42**	.43**	.41**	.68			
7. MX16	.53**	.56**	.35**	.45**	.32**	.70**	.52		
8. VR16	.46**	.75**	.38**	.52**	.14	.76**	.76**	.59	
9. R3D16	.21*	.35**	.33**	.20	.07	.52**	.57**	.26*	.71

Note. Formula (2) from Murphy and Davidshofer (1988) applied to the range-corrected correlations in Table 9 using the reliabilities published in the WAIS-IV technical manual (David Wechsler, 2008) and the initial ICAR validation study (Condon & Revelle, 2014). * indicates $p < .05$. ** indicates $p < .01$

Based on the item content, scores on the Three-Dimensional Rotation (R3D16) items on the ICAR16 were expected to correlate the most highly with measures of Gv on the WAIS-IV. The range and reliability corrected correlation between Gv and R3D (r=.35) was greater in magnitude than correlations with other broad ability scores (Table 11) but was only moderate in effect size and only statistically larger than Gs (z= 2.40, p=.02). Furthermore, correlations were smallest with Gs (r=.06) and negligible in effect size, though their correlation was not significantly smaller than correlations with any other BAC scores except Gv and Gc.

Table 11. Steigler's z Test of Differences Between R3D16 and Respective BAC Scores

Target Correlation (Corrected r)	vs.	Comparison Correlation (Corrected r)	z	sig
Gv (.35)		Gs (.06)	2.40	.02*
Gc (.33)		Gs (.06)	2.03	.04*
Gv (.35)		Gsm (.20)	1.35	.18
Gv (.35)		Gf (.21)	1.23	.22
Gc (.33)		Gsm (.20)	1.22	.22
Gf (.21)		Gs (.06)	1.21	.23
Gc (.33)		Gf (.20)	1.17	.24
Gsm (.20)		Gs (.06)	1.14	.25
Gf (.21)		Gsm (.20)	.225	.82
Gv (.35)		Gc (.33)	.08	.93

Note. Z indicates Steigler's z (1980). * indicates $p < .05$. ** indicates $p < .01$.

Like other matrix tasks, matrix Reasoning (MX16) scores were expected to correlate most highly with measures of Gf on the WAIS-IV, relative to other subtests. Contrary to a priori hypotheses, the MX16 scores correlated most highly with Gv (r=.58), but this correlation was not significantly different from the correlation with Gf (z=.42, p=.68) nor Gsm (z=1.52, p=. 13) (Table 12). The lowest correlation was with Gs, though this was not statistically different from the correlation with Gc (z=.04, p=.97) nor Gsm (z=.69, p=.49). It was also expected that MX16 would correlate the highest with the corresponding Matrix Reasoning subtest on the WAIS-IV, but that was not supported by the data (Appendix B). The range and reliability corrected correlation between MX16 and the WAIS-IV Matrix Reasoning subtest was moderate in magnitude (r=.42, p<.001), but indistinguishable from the correlation between MX16 and Block Design (r=.46, p<.001) and Arithmetic (r=.49, p<.001) respectively. Furthermore, MX16 correlated the highest with the Visual Puzzles subtests on the WAIS-IV (r=.66, p<.001).

Table 12. Steigler's z Test of Differences Between MX16 and Respective BAC Scores

Target Correlation (Corrected r)	vs.	Comparison Correlation (Corrected r)	z	sig
Gv (.56)		Gs (.32)	2.29	.02*
Gf (.53)		Gc (.35)	2.12	.03*
Gv (.56)		Gc (.35)	2.07	.04*
Gf (.53)		Gs (.32)	1.95	.05
Gv (.56)		Gsm (.45)	1.20	.23
Gf (.53)		Gsm (.45)	1.08	.28
Gsm (.45)		Gc (.35)	1.01	.31
Gsm (.45)		Gs (.32)	.88	.38
Gv (.56)		Gf (.53)	.39	.70
Gc (.35)		Gs (.32)	.20	.84

Note. z indicates Steigler's z (1980). * indicates $p < .05$. ** indicates $p < .01$.

Letter-Number Series (LN16) scores were expected to correlate most highly with measures of Gf on the WAIS-IV, which was supported by the data (Table 13). Correlations with Gf were significantly higher than those with all other BAC scores. It was also hypothesized that scores on the LN16 task would correlate the lowest with Gs, however, correlations with Gs were not distinguishable from correlations with any other BAC score except for Gf.

Table 13. Steigler's z Test of Differences Between LN16 and Respective BAC Scores

Target Correlation (Corrected r)	vs.	Comparison Correlation (Corrected r)	z	sig
Gf (.70)		Gc (.42)	3.78	<.01**
Gf (.70)		Gsm (.43)	3.45	<.01**
Gf (.70)		Gs (.41)	3.28	<.01**
Gf (.70)		Gv (.44)	3.22	<.01**
Gv (.44)		Gs (.41)	.31	.76
Gsm (.43)		Gs (.41)	.20	.84
Gv (.44)		Gc (.42)	.16	.88
Gc (.43)		Gs (.41)	.14	.89
Gv (.44)		Gsm (.43)	.11	.91

Note. z indicates Steigler's z (1980). * indicates $p < .05$. ** indicates $p < .01$.

Scores on the Verbal Reasoning (VR16) items were expected to correlate most highly with measures of Gf on the WAIS-IV, relative to other broad abilities, however, this was not supported by the data (Table 14). Scores on the VR16 task correlated most highly with the Gv estimate, and this correlation was significantly higher than correlations with all other broad ability estimates. It was also expected that scores on the VR16 task would demonstrate the lowest correlation with Gs, but these correlations were indistinguishable from the correlation with Gc.

Table 14. Steigler's z Test of Differences Between VR16 and Respective BAC Scores

Target Correlation (Corrected r)	vs.	Comparison Correlation (Corrected r)	z	sig
Gv (.75)		Gs (.14)	6.375	<.01**
Gv (.75)		Gc (.38)	4.247	<.01**
Gv (.75)		Gf (.46)	3.767	<.01**
Gsm (.52)		Gs (.14)	3.541	0
Gf (.46)		Gs (.14)	2.906	0
Gv (.75)		Gsm (.52)	2.814	0
Gc (.38)		Gs (.14)	1.93	.05
Gsm (.52)		Gc (.38)	1.427	.15
Gf (.46)		Gc (.38)	.973	.33

Note. z indicates Steigler's z (1980). * indicates $p < .05$. ** indicates $p < .01$.

4. Discussion

To the author's knowledge, the present study is the first independent validation of the ICAR, and the first to examine the construct validity of the tool using a well-validated and theory-informed measure of cognitive abilities. Data from a convenience sample of 97 university students was used to examine the correlations between scores on the ICAR16 and the WAIS-IV. The WAIS-IV is known to measure up to five broad abilities well based on Cattell-Horn-Carrol theory of intelligence, and composite scores were constructed to estimate the broad abilities measured using the guidelines of Lichtenberger and Kaufman (2012). Correlations between ICAR16 subtest scores and the broad ability composite (BAC) scores were corrected for restriction of range and reliability using published variances and reliabilities from the respective ICAR and WAIS-IV norming samples.

In regard to research question 1, "Is there evidence of convergent validity of the ICAR16 when compared with a well-established, individual cognitive ability battery?", *a priori* hypotheses were largely supported. The correlations between the general factors indicated by the four subtests on the ICAR16 and the four composites scores on the WAIS-IV was large ($r=.94$, $p<.001$) and the model fit the data well (χ^2 (19) = 14.23, $p= .77$; RMSEA < .001, 90% C.I. (.000, .062); CFI = 1.000; TLI = 1.045). Correlations between the respective observed general ability scores on the WAIS-IV, the FSIQ and GAI, and the total score on the ICAR16 were also large, as expected. The uncorrected correlations between the overall scores and the ICAR16 total scores did not reach the hypothesized magnitude of $r=.7$ ($r_{ICAR16,\ FSIQ}= .62$, $p<.001$, $r_{ICAR16,\ GAI}= .61$, $p<.001$), but the range and reliability corrected correlations exceeded this level ($r_{ICAR16,\ FSIQ}= .81$ $p<.001$, $r_{ICAR16,\ GAI}= .81$, $p<.001$).

It was hypothesized the correlation between the ICAR16 total score and the GAI would be higher than that between the ICAR16 total score and the FSIQ, as GAI does not account for scores on measures of processing speed and working memory. Contrary to expectations, the ICAR16 total score correlations between GAI scores and FSIQ scores were not distinguishable from one another. Findings support the ICAR16 as a broad measure of general cognitive ability, but on its own may not be sensitive enough to distinguish between “higher order” cognitive abilities (e.g. fluid reasoning, verbal comprehension) and more basic cognitive functions (e.g. working memory and processing speed).

In regard to research question 2, “Which CHC Constructs are related the ICAR types?”, findings only partially supported *a priori* hypotheses. Evidence from range and reliability corrected correlations suggests that the ICAR subtests are strongly related to fluid reasoning (Gf) and visual-spatial processing (Gv) abilities. As expected, the Letter-Number Series subtest correlated highest with the fluid reasoning estimate, and this correlation was significantly larger than correlations with other abilities. The other ICAR16 subtests also demonstrated moderate to high correlations with Gf, but all three also demonstrated relatively larger correlations with Gv.

Contrary to expectations, the Matrix Reasoning subtest and Verbal Reasoning subtest both correlated highest with measures of Gv. Visual-spatial processing is strongly related to fluid reasoning, and there has been debate among scholars as to whether or not the two are actually distinct constructs (Dombrowski, Canivez, Watkins, & Alexander Beaujean, 2015; Weiss et al., 2013).

For the Matrix Reasoning subtest, the correlations with Gf and Gv were not distinguishable, but nor were they distinguishable from the working memory estimate (Gsm), which is related to but consistently distinct from Gf (Ackerman, Beier, & Boyle, 2005; Engle, Tuholski, Laughlin, & A, 1999; T. L. Harrison, Shipstead, & Engle, 2015; Heitz et al., 2006). At the subtest level, the MX16 correlated the highest with the Visual Puzzles subtest ($r=.66$, $p<.001$) rather than the Matrix Reasoning subtest as expected ($r=.42$, $p<.001$). Some recent research has demonstrated that Gv is more important in matrix tasks than previously believed, and in some groups, equally important as Gf (Waschl et al., 2017). Taken together, findings suggest that the matrix task on the ICAR primarily measures Gv (and perhaps more so than other matrix tasks), though it likely involves multiple abilities. It should also be noted that the reliability of the Matrix task published in the ICAR norming study was only marginally acceptable ($\boldsymbol{\alpha}=.52$), so corrected and uncorrected correlations should be interpreted with caution.

The high correlations between Gv and the Verbal Reasoning subtest are not easily explained. Although Gf and Gv are highly related, the differentiating factor is the quality of visual stimuli, which is completely absent in the VR16 task. The VR16 task correlated highest with scores on the Block Design ($r=.59$, $p<.001$) and Visual Puzzles ($r=.68$, $p<.001$) subtests, which were significantly higher than correlations with the subtest most similar in item content, Arithmetic ($r=.39$, $p<.001$). This finding may result from imprecise representation of the Gv factor by the WAIS-IV subtests; at least one study indicated the Visual Puzzles subtests relies on multiple abilities in a mixed clinical sample (Fallows & Hilsabeck, 2012). Given the inconsistency with theory, the strength of the relation between Gv and the Verbal Reasoning task may be a statistical artefact, and replications are necessary.

Hypotheses about the relations of broad abilities with the Three-Dimensional Rotation subtest were weakly supported by the data. Correlations were the highest with Gv, but these correlations were not significantly different from correlations with any other ability construct aside from Gs. This finding may be due to the difficulty of the R3D16 items. Item information functions from the initial validation study indicated that the discrimination of the R3D16 items peaked at about one standard deviation of latent ability higher than the other item types (Condon & Revelle, 2014). Furthermore, informal behavioral observations from study administrators noted that participants often reported needing much more time to complete these items, and that the items were more difficult. Including R3D items that are similar in difficulty to the WAIS-IV tasks may reveal more distinct relations with specific abilities.

As expected, the ICAR subtests demonstrated the lowest correlations with the processing speed ability estimate, Gs, except for LN16 which demonstrated a negligibly smaller correlation with Gsm. All correlations with Gs were significantly lower than the highest correlations between the respective ICAR tasks and BAC scores, however correlations with Gs were not distinctly lower than correlations with other broad abilities in all cases. Consistent with the results from research question 1, findings suggest the ICAR16 is most related to Gf and Gv abilities but may not be sensitive enough to distinguish between distinct abilities at this sample size.

Overall findings provide initial evidence of convergent and discriminant validity of the ICAR16, with correlations between the ICAR16 subtests being generally higher with estimates of fluid reasoning and visual-spatial processing, and generally lower with estimates of ostensibly indirectly related or unrelated abilities, such as processing speed and working memory. Results should be interpreted with caution, as correlations did not reveal clear distinctions between the constructs measured and suggest that the ICAR16 and/or the sample are not sensitive enough to distinguish between broad abilities. Correlational evidence is considered preliminary to more advanced statistical methods in larger samples that may reveal more well-defined associations with certain broad abilities.

5. Limitations

The present study has several important limitations that challenge the generalizability and interpretability of the findings. First, the heterogeneity of item difficulty may have confounded the correlation comparisons. Some ICAR tasks, such as the Three-Dimensional Rotation items, were much more difficult than the WAIS-IV items, while other tasks, such as the Verbal Reasoning items which involve logic or word problems, can be easily solved by a university-level student with unlimited time.

The generalizability of the results is challenged by the nature of the sample. First, homogeneity in regard to race, ethnicity, education, and socioeconomic status limits the generalizability of the findings. The inclusion of a clinical convenience sample further reduces the generalizability of the findings; although the majority of students who seek assessment services at the university-based center perform in the average to superior range on cognitive testing, psychopathologies such as specific learning disorders, ADHD or mood disorders that are common among this population may influence performance on both the ICAR16 and the WAIS-

IV. Thus, these analyses should be replicated in larger, more diverse samples in order to ensure generalizability.

The most salient limitation of the present analyses is the relatively small sample size. Although adequately powered to detect the observed correlations between subtests, the collected sample was not powered to detect smaller differences between correlations. It is possible that in a larger, more highly powered sample would be sensitive enough to better differentiate between distinct abilities. Similarly, due to the small sample, correlations were corrected for restriction of range and reliability. Although these corrections improve the accuracy of the observed correlations, they call into question results from Steigler's z tests, which are intended for uncorrected Pearson Correlations. Significance tests are not intended for corrected correlations, and conclusions about statistically differences between correlations warrant caution.

Correlational methods can only provide evidence of convergent and discriminant validity; they do not confirm which ability constructs contribute to performance on the respective tasks. Moreover, larger sample sizes are needed to account for multiple statistical comparisons of correlations. Visual inspection of correlations is qualitative in nature, so findings are considered exploratory and provisional. Confirmatory factor analytic approaches are necessary to parse out the effect of method of administration of the respective tests and error (Marsh, 1994; Nussbeck, Eid, & Lischetzke, 2006). The results from this study can be used as preliminary evidence of convergent and discriminant validity of the ICAR16 and verifies that correlations are sufficiently large to conduct factor analytic approaches.

6. Conclusions

The results of the present study provide evidence that the ICAR16 is a valid brief measure of general intelligence, and correlates with general ability estimates on the WAIS-IV at a magnitude similar to other brief measures of intelligence (Floyd, McGrew, & Evans, 2008; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Johnson, te Nijenhuis, & Bouchard, 2008; Salthouse, 2014; Sanders, McIntosh, Dunham, Rothlisberg, & Finch, 2007). The ICAR subtests demonstrated the largest correlations with fluid reasoning and visual spatial processing tasks on the WAIS-IV, and the smallest correlations were with measures of processing speed or working memory. However, many of these correlations were indistinguishable from correlations with other broad ability estimates. Future research should confirm the strength of these relations using larger sample sizes and factor analytic methods.

Limitations aside, the present study offers several strengths as well. It is the first study to examine a public domain measure of cognitive abilities with a well-established, traditional battery. Results provide a theoretical context for the ICAR that previously did not exist, which will spur the development of new and improved items for a wider range of ages and ability levels. The current sample may serve as a linking sample for the norming samples of the ICAR and WAIS-IV respectively and facilitate planned missingness research designs. Finally, findings provide a foundation for future research on the validity of the ICAR test as well as for the use of the ICAR test as a brief, self-administered measure of nonverbal cognitive ability.

7. Implications

The present findings have a number of implications for the use of the ICAR in primary research. First, even though scores on the ICAR correlated with scores on the WAIS-IV at magnitudes similar to those observed between similar tests of abilities, it is notable that the method of administration of the ICAR is categorically different than other brief measures. The ICAR is completely unmonitored and self-administered online; examinees have access to the internet, scratch paper, calculators, and unlimited time to complete the items. Considering the freedom of examinees, it is remarkable that the error due to method of administration did not interfere with subtest correlations to a larger degree. For example, the *Shipley-2* shares many more administration characteristics with the WAIS-IV than the ICAR; it is self-administered but must be monitored by a trained examiner, subtests are time-limited (10 minutes for Vocabulary and 12 minutes for Block Design/Abstraction), and it includes a verbal scale in addition to the nonverbal scale (parinc.com). Despite more administration similarities, Lodge (2012) found the Shipley Composite A to correlate with WAIS-IV FSIQ at a similar magnitude to the ICAR16 total score in a comparable population ($r_{Shipley\text{-}2,FSIQ} = .78$, $p<.001$, , $r_{ICAR16, FSIQ}=.62$ $p<.001$)[1].

Equivalence of public domain tools to commercial measures could translate to substantial savings for researchers who are interested in brief measures of general cognitive ability. Researchers who test a sample of 100 participants with the *Shipley-2* would spend $767 on materials and anywhere from 30 to 250 minutes in administration time, depending on individual or group administration (parinc.com), while the costs of using the ICAR are negligible. The time and money savings of the ICAR compared to a WAIS-IV are even more impressive; the WAIS-IV costs $2,129.65 for materials and an average 900 minutes of administration and scoring time (pearsonassessments.com). In this regard, the ICAR offers obvious advantages, especially for studies in which an estimate of cognitive ability can serve as a covariate (see Appendix F for a review of potential areas of application). The ICAR will also be particularly useful in large-scale studies in which administering individual or even group measures to hundreds of participants is not feasible regardless of budget. However, brief measures like the ICAR are typically not good

[1] Lodge (2012) provided uncorrected Pearson correlation coefficients from a much smaller sample of university students (N=25). Uncorrected correlations from the present study were provided for consistency.

substitutes for traditional batteries for researchers whose primary variables of interest are specific cognitive abilities.

Strengths aside, it is important to highlight the limitations of the ICAR in its current form. Weakness in the psychometric properties of the ICAR item set present a problem for usability of the tool. Like many other brief measures, the internal consistency of the 4-item subtests is low, especially for the Matrix Reasoning items. Only one study to date has examined the measurement invariance of the tool for examinees of different ages and genders (Young, Keith, & Bond, 2019), and no studies have examined invariance across other salient dimensions of identity such as race/ethnicity, nationality, or native language. As the tool is intended for wide-spread international use, further rigorous validation is necessary to ensure the items measure the same constructs equally for all examinees.

One of the major findings of the present study is that the ICAR16 in its present form is not a multidimensional measure of cognitive abilities. It provides a broad estimate of combined fluid and visual spatial reasoning, but it cannot differentiate between the two constructs. This imprecision limits the usefulness of the tool and warrants caution for its use in fields outside of psychology which may not be as familiar with the nuances of cognitive ability measurement. For example, there has historically been some debate over which broad abilities are most representative of general intelligence (see a discussion between Robinson (1999, 2005) and Ashton et al (2001; 2005)), let alone a unified definition for “general intelligence” as a construct more broadly. From this lens, one could argue that the ICAR should be seen as a test of perceptual reasoning specifically and not necessarily a proxy for general intelligence more broadly.

Offering an ostensibly valid intelligence test to the public domain also brings up ethical concerns around who is using the tool and for what purposes. The history of intelligence testing is fraught with controversy and the legacy of oppression in the field should not be underestimated (see Appendix G). Messick (1979) argues that contemporary intelligence test developers have an ethical responsibility to evaluate not only the validity of assessments in terms of their measurement properties, but also the social consequences of their use. One of the advantages of proprietary measures is that test developers are generally trained in the proper use and application of intelligence tests and can control how the tools or validation data are used. For example, the developers of the Woodcock-Johnson Tests of Cognitive Abilities have refused to release information about participant race/ethnicity for certain types of research studies. Researchers also must apply to use the ICAR items and agree to use the items according to standards put forth by the British, American, and German Psychological Societies as well as "participate in this project in a manner other than one which observes the highest standard of ethical conduct" (icar-project.com). Despite these assurances, once the items are published, controlling their use is difficult if not contrary to the spirit of the public domain.

Several papers have been published using the ICAR since 2014, many of which focus on associations between personality and cognitive ability (see Dworak et al., 2020) though some are more polarizing than others. For example, Fuerst and Kirkegaard (2016) "set out to determine whether there was a fairly consistent, positive relationship of racial ancestry with both cognitive and socioeconomic outcomes in the Americas (p. 351)" using the ICAR. Methods and results aside, the goal of the authors is ambiguous and reminiscent of the unsavory past of intelligence testing, and the interpretation and applications of this type of research warrants extreme caution. Because intelligence is often erroneously conflated with human worth in our society, and intelligence testing has many potentially oppressive applications, responsibility falls on research consumers to critically and ethically interpret findings and their implications. Rather than acting as gatekeepers for access to the tool, the research community should instead rigorously evaluate not only the methods, analyses, and design of research produced using the ICAR, but also its guiding theory, goals, and potential applications.

Potential ethical concerns should not deter the use of the ICAR, but rather should be at the forefront of conversations about its applications, advantages, and limitations. Researchers across fields who are interested in measuring cognitive abilities, even as a covariate, should have a strong background in the ethical use of testing instruments. There is a risk of misapplication with any intelligence test, and publishing testing tools in the public domain may increase that risk, but the risks do not outweigh the benefits. Increasing accessibility to research tools presents the opportunity to drive down research costs, increase researcher productivity, promote data sharing, and improve research quality across fields.

Chapter 2: Integrative Analysis

The study of intelligence is well-served by dozens of measures of cognitive abilities, though contemporary research typically relies upon only a few established proprietary batteries (Condon & Revelle, 2014; McGrew, 2009). Despite their extensive research base, these measures pose several challenges for primary researchers interested in the relationship between the various cognitive abilities and other constructs and outcomes. An accessible and affordable method of accurately measuring cognitive ability is important to understand the relation between intelligence and other factors that influence education, health, and well-being, and design individualized interventions in practical settings. The development and validation of a public-domain cognitive ability resource will facilitate the advancement of research in these areas. For a brief review of the literature on several broad areas of research informed by cognitive abilities, and how they could benefit from access to public-domain measurement tools, see Appendix A.

The following chapter provides context for the validation of a public domain measure of cognitive abilities through comparison with a well-established, theory-based assessment. First, the rationale for and limitations of a public-domain measure, and a description of its first incarnation, the International Cognitive Ability Resource (ICAR) is provided. A brief description the Cattell-Horn-Carroll model of cognitive abilities and how they are measured by the WAIS-IV follows. Finally, construct validity of psychological measures and its evaluation in cross-battery assessment is discussed.

Accessible Cognitive Ability Measures for Researchers

In their introduction of the ICAR, Condon and Revelle (2014) presented several reasons the extant toolkit of proprietary cognitive assessments is not ideal for the purposes of primary researchers, including differences in score interpretation, test content, and administrative flexibility needs. First, the authors argued that though the quality of data is essential, primary researchers are typically not concerned with providing feedback to examinees. While examinees in clinical and selection settings have a principal interest in test results and their interpretation, participants in research studies may be motivated by monetary rewards, course credit, or other small incentives (Brase, 2009). Thus, commercial measures that emphasize the interpretive potential of their products are misapplied and inefficient in primary research contexts.

In addition, Condon and Revelle (2014) noted the test content and organization of existing measures is often not optimal for the needs of primary researchers. Most test companies distribute their measures as full kits that include a fixed battery of measures ("Houghton Mifflin Harcourt," 2017; "Kaufman Assessment Battery for Children, Second Edition," 2017; "Wechsler Intelligence Scale for Children®-Fifth Edition," 2017). Almost no batteries contain a sufficient number of tests to measure all of the abilities associated with intelligence (Alfonso, Flanagan, & Radwan, 2005). A researcher interested in measuring a wide variety of cognitive constructs may need to administer tests from several different commercial measures. Alternatively, a researcher interested in only one construct may need to purchase an entire kit, or multiple kits, to access appropriate measures of the construct of interest.

The logistical and practical aspects of administering proprietary measures may pose the most salient barriers to their use in primary research. Practical considerations of accessing proprietary measures include, but are not limited to, the financial cost of the measure and scoring materials (Yates & Taub, 2003), the time needed to administer and score assessments (Camara et al., 2000), requisite training level of administrators (Alfonso, Johnson, Patinella, & Rader, 1998; Mrazik, Janzen, Dombrowski, Barford, & Krawchuk, 2012), and privacy and security of testing and storage spaces (Kaufmann, 2009). Although these considerations may be superficial, they have the potential to burden research progress.

As budgets for research in the social sciences tend to be relatively small compared to other fields (National Science Foundation, 2017), the cost of measures is a substantial consideration for researchers in these areas. The financial cost of proprietary assessments for researchers varies, but most commercial test kits cost hundreds to thousands of dollars (“Houghton Mifflin Harcourt,” 2017, “Pearson Clinical,” 2017, “WPS,” 2017). In addition to the cost of the assessment materials, the cost of trained administrators, appropriate testing spaces, scoring and interpretation time, and clinical supervision inflate the price of administering commercial measures substantially. For example, Yates and Taub (2003) estimated the cost of administering just one WISC-V and a child behavior checklist (CBCL) to be up to $700 per child when all expenses are considered. Though some testing companies offer reduced-rates to qualified research groups, relatively few proprietary measures are freely available to researchers (Condon & Revelle, 2014).

The most widely used freely-available measure of cognitive abilities in psychological research is the Educational Testing Services (ETS) Kit of Factor-Referenced Cognitive Tests. The current version of the kit was published in 1976, and has since been used in hundreds of studies (Babcock & Laguna, 1997; Ekstrom & Bejar, 1990), however, the ETS Kit poses several significant disadvantages to researchers as well. First, though the test is widely used in research, little research has been conducted on the psychometric properties of the measure. In one of the few studies examining the kit, Babcock and Laguna (1997) found a three-factor solution that was inconsistent with the four constructs ETS claimed the kit measured. The authors concluded that the measure should be used with caution in research and suggested it should not serve as the standard against which other tests are compared. No studies to date have compared the ETS kit with a well-established measure of intelligence. In contrast, commercial measures offer an extensive research base, as they are privately validated and have been a focus of independent intelligence researchers as well (e.g. Flanagan & Alfonso, 2016; Keith, Low, Reynolds, Patel, & Ridley, 2010; McGrew, 2009; Weiss et al., 2013). Thus, primary researchers may have to compromise test validity for cost at the expensive of quality research findings.

Direct costs are not the only limiting factor to consider in research; time and training are also imperatives in the implementation of research studies. Common batteries require one-to-one administration for more than one hour, followed by an additional 10 to 20 min for scoring (Camara et al., 2000). Furthermore, standardized administration by a trained professional (typically a masters-level clinician or above) is imperative in order to preserve the validity of the measures and obtain accurate results (Mrazik et al., 2012; Styck & Walsh, 2016). Aside from the costs associated with using trained administrators, the time required to recruit and train such administrators for large samples can be an obstacle in itself. For example, Alfonso and colleagues recommended that students complete five to six administrations before they are competent to administer a valid assessment using the WISC-III (Alfonso et al., 1998). The amount of time to train and supervise administrators is a sizeable obstacle to commercial test use in large studies.

Issues of secure testing and storage space may also be a barrier to the use of proprietary measures in research. Material test kits are physically cumbersome and require secure storage spaces for safe keeping. Federal courts require the protection of psychological tests as a unique methodology, and some states mandate the safeguarding of test materials (Kaufmann, 2009). This means researchers must have locations with adequate security to store materials.

Furthermore, valid administration of most traditional cognitive assessments requires the use of private, distraction-free testing spaces (Strauss, Sherman, & Spreen, 2006). These logistical challenges may increase the likelihood of researchers ignoring cognitive abilities in their work completely.

Group intelligence assessments offer a partial solution to some of the limitations posed by individual tests. Popular group tests such as the Cognitive Abilities Test (CogAT; Lohman et al., 2001), the Cattell Culture-Fair Intelligence Test (CCFIT; Cattell & Cattell, 1960), Raven's Progressive Matrices (RPM; Raven, 1998), the Scholastic Aptitude Test (SAT; Donlon, 1984), among many others, can assess large groups of individuals simultaneously without the need for private test spaces, highly trained administrators, or complex scoring systems (Motta & Joseph, 2000). Group tests of cognitive abilities tend to be relatively more economical than individual assessment (Motta & Joseph, 2000), yet the vast majority are still proprietary measures; primary researchers cannot readily access these measures without going through the proper channels and paying associated fees. Albeit to a lesser extent than individual measures, proprietary group tests present many of the same financial, test security, and logistical concerns to primary researchers as proprietary individual assessments.

One emerging solution to the accessibility problem in cognitive ability research is the National Institute of Health's (NIH) Toolbox. The aim of the NIH common on set of assessment tools that can be used across populations and disciplines to measure cognitive, emotional, sensory, and motor health from ages 3–85 years (Bauer & Zelazo, 2013). The NIH Toolbox is also available at a low cost to researchers, and has been well-validated across child, adult, and clinical populations (Akshoomoff et al., 2014; Hessl et al., 2016; Mungas et al., 2013; Weintraub et al., 2014).

Despite the benefits of the NIH Toolbox for researchers, there are a few areas in which it is lacking in ease of accessibility in comparison to the ICAR. First, albeit at a much lower cost than most commercial measures, the cognitive battery is not completely free of charge (Akshoomoff et al., 2013). Second, trained administrators are still necessary to give the cognitive battery on iPads or on paper, contrary to the ICAR which is self-administered over the web. Finally, the battery is fixed, unlike the ICAR which encourages researchers to create their own items and validate them against the existing battery. Thus, the NIH Toolbox in its current form has more potential for use in clinical settings, while the ICAR remains a more flexible and accessible tool for research.

Barriers to access to cognitive assessment materials in primary research may have wide-spread consequences. Some scholars have suggested that dependence on proprietary measures has slowed scientific progress in the field of psychology (Gambardella & Hall, 2006; Goldberg, 1999; Liao, Armstrong, & Rounds, 2008). Almost two decades ago, Goldberg (1999) proposed that policies and practices of commercial test publishers who prioritize their financial interests over scientific advancement hindered progress in the field of personality psychology. In response, Goldberg developed the International Personality Item Pool (IPIP), an internet-based resource to provide easy access to measures of individual differences and foster collaboration between researchers (Goldberg et al., 2006). The IPIP has been found to be more efficient for use in research settings when compared to some proprietary personality measures, and also demonstrated higher reliability than the most commonly used commercial measures (Hamby, Taylor, Snowden, & Peterson, 2016). Success with a public-domain collection of personality data along with the challenges of using proprietary cognitive measures has encouraged researchers to explore a similar type of resource for cognitive ability measurement.

The International Cognitive Ability Resource (ICAR)

Condon and Revelle (2014) developed the International Cognitive Ability Resource (ICAR) with the goal of providing a free, online assessment tool to researchers in the social sciences and encourage the integration of neuropsychological assessment into medical research and practice. In their initial analysis of the ICAR items, Condon and Revelle (2014) assessed four distinct but related tasks, and provided evidence of sound psychometric properties for the items (Condon & Revelle, 2016). The respective multiple choice item types ask participants to identify: the missing element in a pattern of digit or letter combinations (Letter Number Series, LN, 9 items), the missing shape in a 3x3 array of geometric shapes (Matrix Reasoning, MR-I, 11 items), the correct response to a variety of logic questions and math word problems (Verbal Reasoning, VR, 16 items), and the response choice that is a possible rotation of a stimulus figure (Three-dimensional Rotation 3DR, 24 items) (Condon & Revelle, 2014). All of the items are available for qualified researchers upon request, and the authors published a 16-item sample test, the ICAR16, which was validated concurrently with the longer version.

Condon and Revelle (2014) conducted the initial validation of the ICAR items in 3 studies. The first explored the item characteristics, reliability, and structural properties of the ICAR60 by distributing untimed random sets of 12 to 16 items to an online sample of 96,958 individuals (66% female) ages 14 to 90 from 199 countries. Internal consistency for the item composites on the ICAR16 and the ICAR60 were adequate ($\alpha = .81$; $\omega_{total} = 0.83$) and good ($\alpha = 0.93$; $\omega_{total} = 0.94$), respectively. There was substantial variability in the means and standard deviations of the ICAR items, which suggests that untimed, self-administration of the items online did not lead to uniformly high scores (i.e. cheating). This finding is important since one of the main concerns about the development of a public domain measure is that lack of copyright protection will threaten validity (Goldberg et al., 2006).

One of the stated goals in the development of the ICAR items was to avoid item content that could be readily referenced elsewhere, in order to prevent the temptation to search for answers on the internet in an unproctored setting (Condon & Revelle, 2014). In addition to sufficient variability in responses, the item information functions demonstrated a wide range of item difficulties, which suggests that examinees were unable to look up answers easily. The item information functions for the ICAR16 demonstrated adequate reliability across a range of ability levels, most appropriate within 1.5 standard deviations from the mean item difficulty.

The structural properties of the ICAR16 and the ICAR60 were evaluated in Study 1 as well. An exploratory factor analysis revealed a four-factor solution that fit the data well (RMSEA = 0.014, RMSR = 0.01, TLI = 0.99), with each item type represented by a different factor, and with relatively small cross-loadings (Condon & Revelle, 2014). The item loadings on the primary factors ranged from .2 to .7, and the factor loadings on the general factor were .5 for 3DR, .8 for LN, .8 for VR, and .7 for MR (Condon & Revelle, 2014). Though the four factors are associated directly with the respective item types, the authors did not name what the latent factors supposedly measure. That is, though the item types load onto distinct factors, these factors are more of a reflection of the cohesion of the item types rather than the underlying cognitive abilities involved. Thus, it is unclear which cognitive abilities are being assessed by the items.

Construct validity of the ICAR items was evaluated indirectly by comparison with the *Shipley-2,* a brief, self-administered assessment of crystallized ability and fluid cognitive ability (Shipley, Gruber, & Martin, 2009). The *Shipley-2* measures cognitive ability using three subtests that compose two composite scores: Composite A consists of Vocabulary and Abstraction, and Composite B consists of Vocabulary and Block Patterns. Condon and Revelle (2014) reported that, after correction for restriction of range and reliability, the correlations between the ICAR16 and *Shipley-2* Composites (ICAR16 and Composite A, r=.82; ICAR16 and Composite B, r=.81), were similar to those of the *Shipley-2* composites and other tests of cognitive abilities.

Though psychometric data has been reviewed for the *Shipley-2* and deemed adequate (Kaya, Delen, & Bulut, 2012), it does not have a strong independent research base relative to more widely-used measures of intelligence (Reynolds et al., 2016). The construct validity of the *Shipley-2* has been called into question in recent work (Beaujean et al., 2017; Reynolds et al., 2016). Based on a comparative analysis of the *Shipley-2* and the WISC-IV, Reynolds and colleagues (2016) concluded that the *Shipley-2* composites should not be interpreted as strong indicators of psychometric *g* for children and adolescents. Beaujean and colleagues (2017) also advised caution with the *Shipley-2* after finding that the Block Patterns subtest did not measure a unitary construct. Aside from concerns about the reference measure, Condon and Revelle (2014) only provided correlations between the ICAR16 total score and the *Shipley-2* Composites scores. They did not report correlations between specific ICAR item types and *Shipley-2* subtest scores, nor did they report variance explained by a general factor. Thus, further evidence of the construct validity of the ICAR16 is needed.

A comparison study of the ICAR16 with a well-established measure of intelligence grounded in CHC theory is necessary in order to attain a level of confidence in the appropriate uses for the tool (Woodcock, 1990). Cross-battery analysis of data will help elucidate the relationship between the ICAR16 item types and various cognitive abilities and psychometric *g* (Reynolds, Keith, Flanagan, & Alfonso, 2013). Though comparison with a well-studied measure is necessary, it is important to note that the nature of the ICAR is different than traditional measures by design, and thus limitations are inherent in its use.

A brief, online, self-administered, and untimed test is less precise and exhaustive than traditional comprehensive assessment (Kaufman & Kaufman, 2001). The ICAR is not intended for clinical or diagnostic purposes (Condon & Revelle, 2014). The ICAR lacks both standardized administration and interpretable results in a clinical setting. Furthermore, by design, the system does not include behavioral observations from trained clinicians, which is essential to clinical evaluation and decision-making (Oakland, Glutting, & Watkins, 2005; Sattler, 2008). Conversely, the limitations of the ICAR to the clinician present as strengths to the primary researcher--it is time and cost efficient, relatively easy to use for both researchers and participants, and facilitates efficient data collection and analysis.

From the perspective of the primary researcher, the main limitation of the current version of the ICAR items is the paucity of independent research using the measure. In order for the measure to be used with confidence, the constructs measured by the ICAR must be understood in the context of current theories. The following section briefly discusses the current prevailing theory, the Cattell-Horn-Carroll (CHC) theory of cognitive abilities, and how the associated constructs are measured by the WAIS-IV.

Contemporary Intelligence Theory

Despite centuries of interest in the nature and measurement of human cognitive abilities, there continues to be no consensus on a unified definition of intelligence (Legg & Hutter, 2007; Neisser et al., 1996; Sternberg & Detterman, 1986). The historical context of the development of theories of intelligence is extensive and wrought with controversy, and thus is important to bear in mind throughout the development and validation of new intelligence measures. A brief comment on the historical and cultural context of these theories is provided in Appendix G.

Though there are several working theories in the literature (see Das, Naglieri, & Kirby, 1994; Gardner, 1987; Johnson & Bouchard, 2005), Cattell-Horn-Carroll theory has come to be the most popular and well-researched theory of cognitive abilities today (Flanagan & Harrison, 2012; Keith & Reynolds, 2010; McGrew, 2005, 2009), and thus is the focus of the remainder of this section.

Cattell-Horn-Carroll Theory

In a 1957 letter to John Carroll, Raymond Cattell compared the creation of a standardized taxonomy of cognitive abilities in the field of psychology to the fixing of the standard meter in physics, or the calibration of atomic weights in chemistry (McGrew, 2009). The field's own standardization began to emerge in the literature decades later, when Carroll (1993) published *Human Cognitive Abilities: A Survey of Factor-Analytic Studies.* Carroll's seminal meta-analysis provided the first comprehensive, empirically-based, systematic organization of the structure of human cognitive abilities in a unified framework (McGrew, 2009). Carroll´s model represents a hierarchical, three-stratum organization of human abilities based on factor analytic research, with stratum I consisting of narrow abilities, stratum II consisting of broad abilities, and stratum III consisting of a single general ability factor, *g* (Carroll, 1993).

Several decades prior, the Cattell-Horn's Gf-Gc Theory used second-order factor analysis to identify two broad intelligence factors: fluid ability (Gf), which represented biologically based reasoning abilities, and crystallized intelligence (Gc), which represented acquired knowledge (Horn & Cattell, 1966). The theory evolved to incorporate a more diverse range of broad abilities to adequately account for the breadth of human abilities, though never adopted the third-order general intelligence factor (Horn & Blankson, 2005). The presence or absence of a general factor of intelligence remained the main distinction between Cattell-Horn Gf-Gc theory and Carroll's three-stratum theory; otherwise, the theories share more in common than not (Alfonso et al., 2005). Because of similarities between the two models, the Cattell-Horn Gf-Gc theory was integrated with Carroll's three-stratum theory to form what is now the most widely accepted model of human intelligence, Cattell-Horn-Carroll (CHC) theory (Alfonso et al., 2005; Keith & Reynolds, 2010; McGrew, 2009; W. Schneider & McGrew, 2012).

The broad abilities presented in stratum II of CHC theory are important in understanding the influence of cognitive abilities and their measurement. Several studies have provided evidence that the influence of *g* on achievement is mediated through the broad abilities, and that some broad abilities influence achievement over and above the effect of *g* (Floyd, McGrew, & Evans, 2008; Keith, 1999; Taub et al., 2008; Vanderwood, McGrew, Flanagan, & Keith, 2002). McGrew (2009) describes 16 broad abilities involved in CHC model, six of which are "tentatively defined" due to their sensorial/kinesthetic rather than purely cognitive nature (Danthiir, Roberts, Pallier, & Stankov, 2001; McGrew, 2009). Most practical applications of CHC theory address up to nine broad abilities: fluid intelligence (Gf), crystallized intelligence or verbal-comprehension (Gc), visual-spatial ability (Gv), auditory processing (Ga), short-term memory (Gsm), long-term retrieval (Glr), processing speed (Gs), quantitative knowledge (Gq), reading and writing Grw) (McGrew, 2009). For descriptions of all nine of these broad abilities, see Appendix G.

Currently, no single cognitive battery measures the full breadth of the CHC abilities (Alfonso et al., 2005). Different batteries serve distinct purposes and measure different constructs. Because of the self-administered nature of the ICAR, certain abilities such as writing, auditory cognition, processing speed, and long-term memory are not feasibly assessed. Thus, only broad abilities that are reasonably assessed by the ICAR and the WAIS-IV, including Gf, Gsm, Gv, Gc, and Gs, are relevant to the scope of this paper. CHC-based perspectives on the broad abilities and common approaches to their measurement is briefly discussed in the following section.

Fluid Intelligence (Gf). McGrew (2009) defines Gf as the use of deliberate and controlled mental operations to solve novel problems. Gf is of particular interest to test developers because of its strong relation to g. There is some empirical and theoretical evidence of the equivalence of Gf to *g* (Gustafsson, 1984, 1988, 2002; Reynolds & Keith, 2007; Undheim & Gustafsson, 1987), but other research has conflicted with this notion as well (Blair, 2006; Gignac, 2014, 2015). Disagreement over the interchangeability of Gf and *g* may be reconciled to an extent by Cattell's Investment theory (1987). Investment theory asserts that Gf reflects a neurobiological ability to perceive abstract and concrete relations, and is responsible in part for the acquisition of knowledge and skills (Cattell, 1987; Kvist & Gustafsson, 2008). Evidence from neuroimaging studies suggests that Gf is more biologically rooted in the brain relative to other broad abilities (Bigler et al., 1995; Gong et al., 2005). Furthermore, Kvist and Gustafsson (2008) found that *g* and Gf were equivalent within populations that had equal opportunities to learn, but demonstrated a weaker association in heterogeneously educated populations. Through this perspective, Gf can be viewed as the aptitude to learn within a given environment, which drives the development of other areas of ability.

According to McGrew and Evans (2004), fluid intelligence is comprised of five stratum I abilities: General Sequential Reasoning (i.e. the ability to apply a given set of rules to solve a problem; RG), Induction (i.e. the ability to discover underlying rules from a given set of observations; I), Quantitative Reasoning (i.e. the ability to use reasoning in regards to numerical relations; RQ), Piagetian Reasoning (i.e. the ability to use apply cognitive concepts defined by Piagetian developmental theory; RP), and Speed of Reasoning (i.e. the speed with which one can apply reasoning skills; RE). Of these narrow abilities, quantitative reasoning (RQ), induction (I), and deduction (RG) are most commonly measured in cognitive assessments (Carroll, 1993). Wilhelm (2005) observed that deductive reasoning tends to be measured by verbal stimuli, while inductive reasoning tends to be measured by figural-spatial stimuli, and quantitative reasoning tends to be measured using numerical stimuli. Some researchers have suggested that depending on the nature of the task, tests of these narrow abilities may be influenced by other broad abilities, like Gv or Gc (Waschl, Nettelbeck, Jackson, & Burns, 2016; Wilhelm, 2005). Still, Carroll (1993) found that measures of inductive reasoning tend to be the most representative of Gf.

Both Cattell (1971)and Carroll (1993) observed that the best tests of general intelligence involve tests of Gf, and most short-form measures of intelligence focus on measuring Gf (e.g. Johnsen, 2017; Kaufman & Kaufman, 1990; Shipley, 2009; Wechsler, 1999). Some well-established and widely-used propriety tests use a single task, typically inductive reasoning tasks such as matrices or pattern recognition, to measure Gf (e.g. Cattell & Cattell, 1960; Raven, 1998). Most long-form proprietary cognitive batteries measure Gf through multiple non-verbal tests or tests of abstraction in order to avoid task-specific variance (e.g. Elliott, 2007; Schrank et al., 2014; Wechsler, 2014). These tests typically involve at least one matrices or pattern-recognition-type task. The content of the ICAR items (i.e. pattern recognition through a variety of stimuli) suggests that Gf is the primary focus of the test.

Short-term Memory (Gsm). McGrew (2009) defines Gsm as the ability to apprehend and maintain awareness of a limited number of elements of information in the immediate situation. The Gsm broad ability is comprised by two distinct narrow abilities, Memory Span (i.e. the ability to attend to and immediately recall presented information; MS), and Working Memory (i.e. the ability to mentally hold and manipulate information while performing some mental operation; MW) (McGrew & Evans, 2004). MS is thought to demand less mental resources than MW, which has been supported by empirical data (Kail & Hall, 2001; Reynolds, 1997). In fact, MW is thought to be an essential influence on Gf, and contributes substantially to general intelligence (Gignac, 2014). Unlike MS, psychometric evidence suggests MW measures represent a factorially complex mixture of abilities rather than a cohesive latent factor (Carroll, 1993; McGrew & Evans, 2004). The MW is included among the narrow abilities because of the theoretical and experimental evidence of its existence (Kyllonen & Dennis, 1996; McGrew & Evans, 2004). McGrew and Evans (2004) integrate components of the most-widely accepted theoretical model of working memory, the Baddeley model, in the CHC definition of MW (Leffard et al., 2006). These components include the phonological loop, which processes verbal and auditory information, the visual spatial sketchpad, which processes visual and spatial information, and the central executive system that directs resources between the two systems (Baddeley, 2001; McGrew & Evans, 2004).

Different types of tasks are thought to measure the various components of short term memory. Simple span tasks, such as repeating a series of digits in order, are typically thought to be good measures of MS (Leffard et al., 2006; C. R. Reynolds, 1997). These types of tasks are not measured on the ICAR in its current version but could be measured by automated computer systems. MW by nature is more complex to measure. Tasks that require some mental transformation of information are typically included in working memory composites, such as sequencing a series of digits or letters or reciting the series backward (Leffard et al., 2006).

Mental problem solving of verbally presented problems may measure Gsm. Some four-factor models of the Wechsler scales suggest that the Arithmetic task loads onto the Gsm factor (Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Wechsler, 2008; Weiss et al., 2013), however, at least one model found Arithmetic is a better measure of Quantitative Knowledge (Gq) and Processing Speed (Gs) than Gsm (Phelps, McGrew, Knopik, & Ford, 2005), and many other models have found it to be a better indicator of Gf (Keith et al., 2006; Niileksela et al., 2013; Weiss et al., 2013). Keith and colleagues (2006) suggested that the task calls upon a diverse set of abilities (Keith & Reynolds, 2010). It is unclear the extent to which the working memory skills involved in the Arithmetic are method-specific. That is, it is unlikely the role of working memory would be if the same problems from the Arithmetic subtests were self-administered (like in the ICAR) rather than administered orally (like in the WAIS-IV), as this eliminates the need for mental calculations.

Visual-Spatial Ability. McGrew and Evans (2004) define Gv as the ability to generate, store, retrieve, and transform visual images and tactile sensations. Gv abilities are included in almost every model of intelligence, yet their predictive power in terms of achievement outcomes tends to be overshadowed by other abilities such as Gf and Gc (Lohman, 1996). Gv represents a collection of ten narrow abilities, that involve the mental manipulation of objects (VZ, SR), ability to identify visual forms among distracting visual information (CS, CF, SS), mental estimation of spatial distances (LE), or the ability to perceive and depict abstract visual information (IL, PM, IM). Though Gv abilities are not strong predictors of academic success, some have argued they are less reliable measures of psychometric g (Lohman, 1996; Spearman

& Jones, 1950). Spatial ability measures are gaining interest as indicators of creativity, success in gifted identification, and aptitude in STEM careers (Andersen, 2014; Wai, Lubinski, & Benbow, 2009).

Lohman (1996) identifies four different methods for measuring spatial abilities: performance tests, paper-and-pencil tests, verbal tests, and dynamic computer based tests. Performance tests such as block manipulation, and paper-and-pencil tests, such as paper folding tasks, are among the earliest measures of spatial intelligence and still used in contemporary cognitive batteries (Lohman, 1996). Factor-analyses of large data sets from paper-and-pencil measures have revealed five distinct factors measured by these types of tests (Carroll, 1993; Lohman, 1979), and thus it is recommended that multiple tests are used to measure the construct.

Verbal-Comprehension Knowledge (Crystallized Intelligence) (Gc). Horn described Gc as cultural knowledge acquired through the process of acculturation (Horn, 1991). According to McGrew and Evans (2004), Gc is primarily language-based declarative and procedural knowledge acquired through the application of other abilities in educational and other experiences. Narrow abilities include areas such as language development (LD), lexical knowledge (VL), listening ability (LS), content knowledge (K0, K2), communication skills (CM, OP, MY), and foreign language skills (KL, LA). Most major tests of intellectual ability contain some verbal assessment, and verbal comprehension is thought to be an important indicator of g (Sternberg & Powell, 1983). In fact, verbal subtests have been found to be one of the best indicators of general intelligence (Gignac, 2006; Robinson, 1999), and Gc is the only ability that does not deteriorate with age (Salthouse, 2009). Verbal subtests are often thought to decrease the cultural fairness of tests, though cross-cultural research has demonstrated that non-verbal tests can be culturally biased as well (Rosselli & Ardila, 2003). Despite lingering cultural concerns, verbal subtests remain a staple of contemporary intelligence batteries.

Gc is most often directly measured by tests of vocabulary, reading comprehension, and cultural information (Sternberg & Powell, 1983). Given the verbal nature of many tests, there is evidence Gc may also indirectly influence performance on subtests that are not intended to measure verbal comprehension explicitly, such as Arithmetic (Keith et al., 2006; Weiss et al., 2013). Measuring verbal comprehension becomes more challenging in self-administered online formats since vocabulary and cultural information is easily searched on the internet. This should be considered in the development and evaluation of online public domain assessments of Gc.

Processing Speed (Gs). McGrew and Evans (2004) define Gs as the ability to automatically and fluently perform simple or over-learned cognitive tasks, especially when a high degree of focused attention is required. The narrow abilities associated with Gs include perceptual Speed (P), Number Facility (N), Reading Speed (RS), Writing Speed (WS) and Rate-of-Test-Taking (R9). Gs is considered a more discrete domain of cognitive functioning relative to other broad abilities that are considered to be "higher-order" such as Gf and Gc (Borghesani et al., 2013; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Wechsler, 2008). Although Gs tends to demonstrate a lesser relation with g relative to other broad abilities, there is evidence that Gs is a neurologically-based foundation upon which more complex abilities are developed (Fry & Hale, 1996; Kail, 2000). Along with Gsm, Gs tends to be substantially impaired among neurologically compromised individuals, to the extent that estimates of general intelligence are often calculated without these scores for clinical populations (A. G. Harrison, DeLisle, & Parker, 2008). It follows that discrepancies between Gs and higher order ability scores can be indications of neuropsychopathology (Lichtenberger & Kaufman, 2009).

Measures of Gs involve different types of rote perceptual, motor, or cognitive tasks, and various degrees of motor speed and agility (Kail, 2000). Although Gs may indirectly influence performance on a range of different types of intelligence tasks, direct measures of processing speed are timed tasks (Kail, 2000; Vernon, 1983). It is unreasonable to expect that a self-administered untimed task is a direct measure of Gs. Thus, the subtests that measure Gs on the WAIS-IV are used as a point of discriminant validity of the ICAR items.

Wechsler Adult Intelligence Scales--Fourth Edition.

The Wechsler scales are among the most well-established and commonly used measures of cognitive abilities (Alfonso et al., 2005; Lichtenberger & Kaufman, 2009). Since the release of the fourth edition of the Wechsler Adult Intelligence Scale, a substantial amount of research has been conducted on its factor structure, including several studies grounded in CHC theory (Benson et al., 2010; Niileksela et al., 2013; Schneider & McGrew, 2012; Ward, Bergman, & Hebert, 2012; Weiss et al., 2013). The *WAIS-IV Technical and Interpretive Manual* reports a number of confirmatory factor analyses (CFA) that support a four-factor structure of the test norming data. The four index scores, the Verbal Comprehension Index (VCI), the Perceptual Reasoning Index (PRI), the Processing Speed Index (PSI), and the Working Memory Index (WMI), are derived from these factors (Wechsler, 2008). At least two independent analyses of the WAIS-IV have also supported the four-factor structure, including Canivez and Watkins (2010) and Ward, Bergman, and Herbert (2012), but through divergent perspectives. Ward and colleagues used cognitive theory to help specify and interpret their models, while Canivez and Watkins argued for a bi-factor approach in which the second-order g-factor is the primary focus of interpretation. Four-factor models have some empirical support but deviate slightly from CHC theory in that they tend to combine the broad abilities of Gf and Gv.

Perhaps the most notable study of the structure of the WAIS-IV is by Weiss and colleagues (2013), which was featured along with eight commentaries in a special issue of the *Journal of Psychoeducational Assessment* dedicated the fourth editions of the Wechsler Scales, Weiss and colleagues (2013) found that based on all 15 WAIS-IV subtests, both four- and five-factor models fit the data, but the five-factor model provided a better fit. The five-factor model differentiated the combined Gf/Gv factor ("Perceptual Reasoning") cited in the WAIS-IV technical manual into separate Gf and Gv latent factors, which fit the data better and was more consistent with CHC theory (Weiss et al., 2013). In the five-factor model, Block Design, Picture Completion, and Visual Puzzles loaded on the Gv factor while Matrix Reasoning, Figure Weights, and Arithmetic loaded on to the Gf factor. This interpretation is supported by the work of Benson and colleagues (2010), who also found that a five-factor CHC-based structure fit the data better than the four-factor structure offered by the test developers.

Weiss and colleagues' models provide evidence that the 15 subtest WAIS-IV is a valid measure of five CHC broad abilities, however, most examiners do not administer all of these subtests in practice, and they are not administered to examinees aged 70 to 90. To remedy this problem, Niileksela and colleagues (2013) found a CHC-based five-factor alternative solution using just the ten core subtests. The authors achieved this solution by allowing the three-digit span tasks (digits forward, digits backward, and digit sequencing) to load as three separate indicators on the Gsm latent factor while Arithmetic and Matrix Reasoning loaded on the Gf factor. This solution fit the data well and provides evidence of construct validity for core battery of WAIS-IV subtests according to CHC theory (Niileksela et al., 2013). The separation of the digits forward, digits backward, and digit sequencing subtasks has been supported by a number of studies using factor-analytic and brain imaging techniques (Gerton et al., 2004; Griffin &

Heffernan, 1983; Reynolds, 1997; Unsworth & Engle, 2007). Otherwise, the model was fairly consistent with the five-factor model by Weiss and colleagues (2013).

Lichtenberger and Kaufman (2009, 2012) offer an interpretive system for the full battery and the 10 subtest core battery based on a five-factor model by Keith (2009), which is congruent with the predominant five-factor models in the literature (i.e. Niileksela et al., 2013, and Weiss et al., 2013). The interpretive system offers two options for calculating estimates of the broad ability factor scores, referred to as Broad Ability Composite (BAC) scores. Both versions calculate broad ability estimates by summing the scores on two related subtests, respectively, however, the Keith Five-Factor model is based on the use of Letter-Number Sequencing and Figure Weights, and the Core Five-Factor model excludes those two subtests and relies exclusively on the core battery (Table 6). This system of interpretation allows for association of subtest data with CHC constructs when factor analysis of large sets of data is not possible. For this reason, the Core Five-Factor Model interpretation was used to associate subtest data with CHC broad abilities in subsequent analyses.

Summary. The Cattell-Horn-Carroll theory of cognitive abilities is the best researched and most widely accepted model of human intelligence in the literature today. Most contemporary tests of cognitive abilities are guided by CHC theory. The broad abilities in stratum II of Carroll's three stratum model are of particular importance to understanding the influence of intelligence. Though each broad ability plays an important role, five broad abilities including Gf, Gv, Gsm, Gc, and Gs are consistently measured on test batteries and are relevant to the cross-battery examination of the WAIS-IV and the ICAR. Research suggests the WAIS-IV measures these five broad abilities well. Estimates of the CHC broad abilities from individual subtest scores can be computed using the interpretive system presented by Lichtenberger and

Kaufman (2012) when factor scores cannot be calculated. Taken together, the WAIS-IV provides a reasonable CHC-based reference with which to compare the ICAR.

Evaluating Construct Validity of Cognitive Assessments

A large body of work has been dedicated to determining what constitutes evidence of construct validity in psychological measurement (Cronbach, 1988, 1989; Foster & Cone, 1995; Messick, 1979). In simple terms, the construct validity of a test is the extent to which a test measures what it is intended to measure (Anastasi, 1968). One approach to investigating the constructs measured by a new test is to compare it to another test with well-known properties (Woodcock, 1990). In cross-battery comparisons, construct validity is supported by evidence of convergent validity and discriminant validity, that is, tests demonstrate a strong relation to other tests of the same construct and are unrelated to tests of distinct constructs (Campbell & Fiske, 1959). In 1959, Campbell and Fiske published a seminal article that presented one of the earliest systematic methods for evaluating convergent and divergent validity of psychological measures: the multitrait multimethod matrix (MTMM) analysis. In order to demonstrate evidence of convergent and divergent validity in an MTMM analysis, different measures of the same trait (monotrait, heteromethod) should demonstrate high correlations, while correlations between measures of distinct traits (heterotrait, monomethod, and heterotrait, heteromethod) should be relatively lower (Campbell & Fiske, 1959).

Despite the massive influence of the work of Campbell and Fiske in the development of psychometric standards in psychology (Sternberg, 1992), numerous authors have identified limitations to their MTMM analytic method (see Schmitt & Stults, 1986), including Campbell and Fiske themselves (Fiske & Campbell, 1992). One of the primary problems with the method is that it is qualitative in nature and based on a visual comparison of patterns of correlations, thus the degree to which convergence/divergence criteria are met cannot be quantified (Lowe & Ryan-Wenger, 1992). Method and error variance are not feasibly separated via visual inspection, thus the relative contributions of error and method variance are unclear (Campbell & O'Connell, 1982).

Advances in technology over the last 50 years have provided alternatives to overcome the limitations posed by traditional MTMM methods. Confirmatory factor analysis (CFA) of a MTMM tests if the hypothesized common trait and method factor structure is supported by the variance observed in the matrix (Wothke, 1996). This method is preferred for a number of reasons summarized by Lowe and Ryan-Wenger (1992). First, the model is able to estimate the trait and method factor loading, the interrelations, and the random errors for each of the variables. Second, the CFA approach is able to statistically test if the theoretical model is plausible given the observed data. Finally, the CFA is able to determine the relative contributions of the trait and method components and remove random error by testing inferences on latent variables rather than observed variables. In this approach, convergent validity is inferred from the loadings of the observed variables on to their respective trait factors, while discriminant validity is supported by near-zero interrelations among distinct latent trait variables (Lowe & Ryan-Wenger, 1992).

Though the advantages of CFA approaches to MMTM evaluation are compelling from a methodological perspective, there are several limitations which make these methods difficult to implement in practice. In order for the model to be identified, a minimum of three traits must be measured by three distinct methods. Additional measures increase study expense, time, and subject demand (Lowe & Ryan-Wenger, 1992). In addition, CFA approaches to MTMM require large sample sizes for even minimally complex models ($N \sim 250$) (Nussbeck, Eid, & Lischetzke, 2006). Collecting samples of this magnitude in most applied settings is impractical within a reasonable timeline due to the limitations of the extant cognitive assessment toolkit outlined in the Integrative Analysis section of this paper. For these reasons, as well as sample quality, the vast majority of intelligence researchers use the nationally normed standardization samples of commercial tests products (McGrew & Wendling, 2010). Standardization samples present many advantages for generalizability but the test companies who sponsor the data collection determine which assessments are administered, which presents a challenge to cross-battery analysis with independent measures.

Planned missingness (PM) designs that use small samples of researcher-collected, full-case data to link large, representative standardization samples may be an alternative to independently collecting large amounts of data (Garnier-Villarreal, Rhemtulla, & Little, 2014; Graham, Taylor, Olchowski, & Cumsille, 2006; Little & Rhemtulla, 2013). In the two-method measurement PM design, inexpensive, less valid measures and expensive, valid measures of the same construct are used simultaneously, with a small subsample of full cases (expensive and inexpensive measures) used to estimate missing information associated with a larger sample of partial cases (inexpensive measures only)(Graham et al., 2006). Though PM designs offer many advantages, several practical considerations must be considered before their implementation.

First, though relatively fewer cases are needed for PM designs, a substantial number of full cases (N~100) are still necessary to ensure model convergence in most cases (Jia et al., 2014). Even with a sufficient number of full cases, as in most confirmatory factor analysis, theory-guided models are specified a *priori* (Keith & Reynolds, 2012). In situations in which theoretical structure of measures are unclear, CFA model specifications are based on face validity and speculation alone. For example, the ICAR developers did not offer theoretical explanations of the factors extracted from the data, so it is unclear which traits the test intends to measure and how many. Moreover, it is recommended that researchers collect small amounts to data to determine *a priori* if there are sufficiently high correlations between the measures to maximize the utility of PM two-method measurement designs (Olchowski, 2008). For small samples traditional methods are more feasible, and are still often implemented in healthcare research (Streiner, Norman, & Cairney, 2015). Correlational studies may serve as an intermediate step towards advanced statistical approaches, as long as the limitations of these methods are considered.

Correlational analyses may be used to find preliminary evidence of construct validity, not confirm its existence. According to Cronbach (1989), “the Campbell and Fiske correlational check is not subtle. A substantial correlation of Trait 1 with Trait 2 does not make their distinction untenable; rather, the correlation puts the advocate under pressure to create conditions under which the variables pull apart” (pg. 154). Therefore, substantial and significant correlations between instruments designed to assess the same constructs are a necessary but not sufficient requirement to establish the validity of any psychological test (Fiske, 1971).

Conclusion

The extant tool kit of cognitive assessments lacks a validated instrument that is freely available to researchers across a variety of fields. Cognitive ability is an important construct that significantly predicts educational achievement, professional success, health outcomes, and well-being (e.g. David Batty, Deary, & Gottfredson, 2007; Lager, Melin, Hemmingsson, & Sörberg Wallin, 2017; Roth et al., 2015; Schmidt & Hunter, 2004), yet it is often overlooked in primary research. Existing proprietary measures, while thoroughly validated and well-suited to clinical use, are not ideal for use in many primary research settings and may impede the measurement of cognitive ability constructs in relevant areas of research. Access to cognitive ability measurement tools in the public domain will help facilitate research progress across the scientific community.

The International Cognitive Ability Resource (ICAR) is the first measure of its kind to be available in the public domain(Condon & Revelle, 2014). The ICAR was validated on a large sample of participants who completed the measure online, and demonstrates initial evidence of strong psychometric properties(Condon & Revelle, 2014). One of the main barriers to the usefulness of the ICAR in research is that no research to date has been conducted on its construct validity using a well-established, theoretically-guided cognitive battery.

Cattell-Horn-Carroll (CHC) theory is the most widely accepted model of human intelligence and serves as the common language for which intelligence researchers can communicate results (Alfonso et al., 2005; Keith & Reynolds, 2010, 2010; McGrew, 2009; Schneider & McGrew, 2012). Most contemporary tests of intelligence are informed by CHC theory, including the WAIS-IV. The psychometric and theoretical properties of the WAIS-IV are well-established, and it consequently offers a useful point of reference to establish convergent and discriminant validity with the ICAR.

Appendix A: Normality Assumption Checks

Examine histograms of WAIS-IV subtest scaled scores.

```
library(tidyverse)
library(dplyr)
library(tidyr)
standardized_subtests<- select(WAISICAR, c("AR_SS", "MR_SS", "BD_SS", "VP_SS", "DS_SS",
"DS_F_SS", "DS_B_SS", "DS_S_SS", "VC_SS", "SI_SS","IN_SS", "SS_SS", "CD_SS") )


colnames(standardized_subtests)<-c("Arithmetic", "Matrix Reasoning", "Block Design", "Visual
Puzzles", "Digit Span Total Score", "Digit Span Forward", "Digit Span Backward", "Digit Span
Sequencing", "Vocabulary", "Similarities", "Information","Symbol Search", "Coding")

graph_wais_data_gathered<- standardized_subtests %>% select(c("Arithmetic", "Matrix
Reasoning", "Block Design", "Visual Puzzles", "Digit Span Total Score", "Vocabulary",
"Similarities", "Information","Symbol Search", "Coding")) %>% gather(key=Test, value=Score)

ggplot(graph_wais_data_gathered) + aes(x=as.numeric(Score)) + geom_histogram(bins =19,
fill = '#000000') + facet_wrap(~Test, ncol=5, nrow=2) + labs(x ='WAIS-IV Subtest Scaled
Score', y = 'Number of Participants')  + coord_cartesian(xlim=c(0, 19)) + theme_minimal()
```

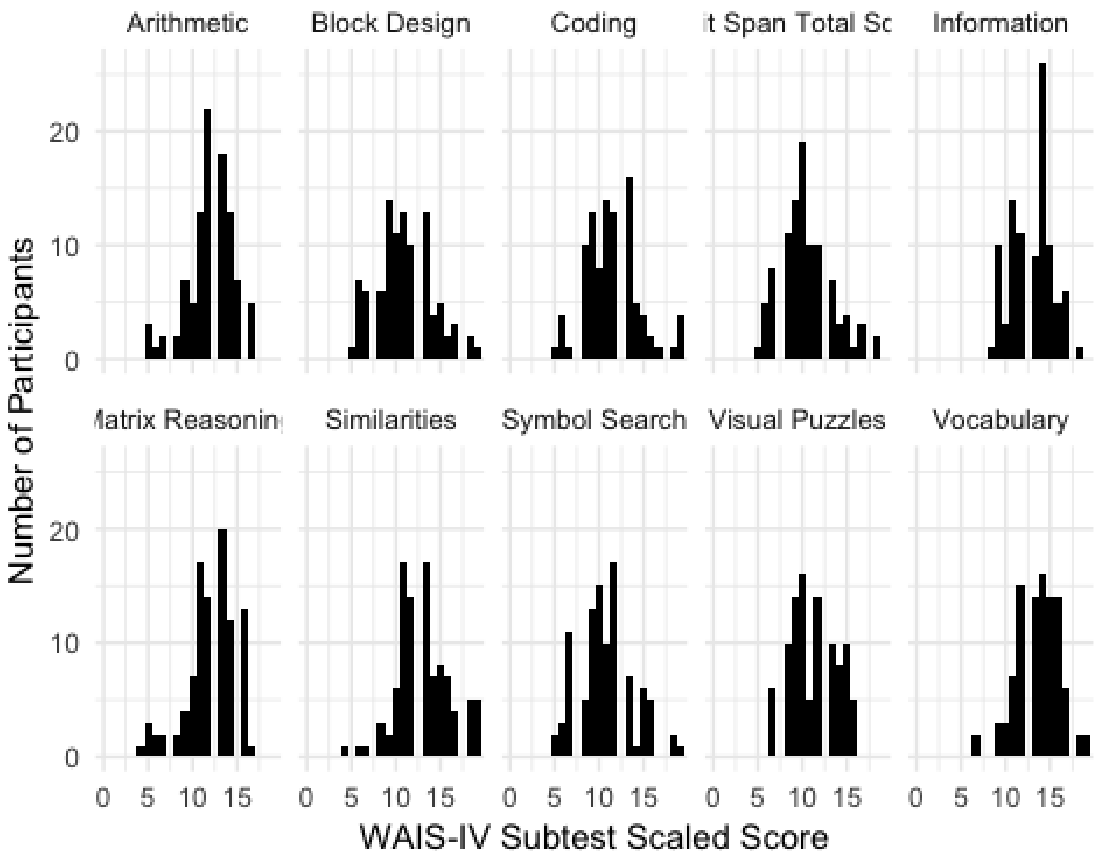


Figure 4. Histograms of WAIS-IV Subtest Distribution

```
graph_icar_data<- WAISICAR %>% select(c(LN16, MX16, VR16, R3D16))

colnames(graph_icar_data)<-c("Letter-Number Series", "Matrix Reasoning", "Verbal Reasoning", "3D-Rotation")


graph_icar_data<-graph_icar_data %>% gather(key=Test, value=Score)

ggplot(graph_icar_data) + aes(x=Score) + geom_histogram(bins =4, fill = '#000000') + facet_wrap(~Test, nrow=2) + labs(x = 'ICAR Subtest Score', y = 'Number of Participants')  + coord_cartesian(xlim=c(0, 4)) + theme_minimal()
```

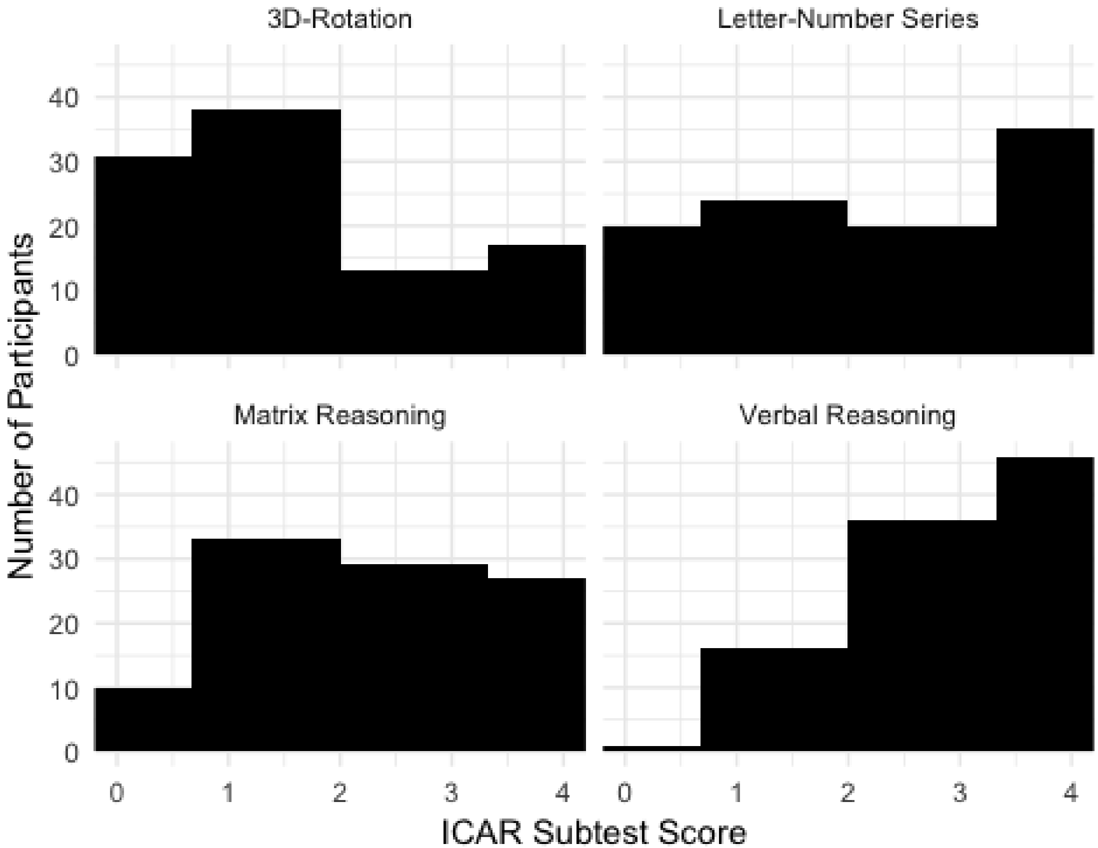


Figure 5. Histograms of ICAR16 Subtest Distributions

```
ggplot(WAISICAR) + aes(x=FSIQ) + geom_histogram(bins =60, fill = '#000000')  + labs(x = 'FSIQ Standard Score', y = 'Number of Participants')  + coord_cartesian(xlim=c(70, 145 )) + theme_minimal()

## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

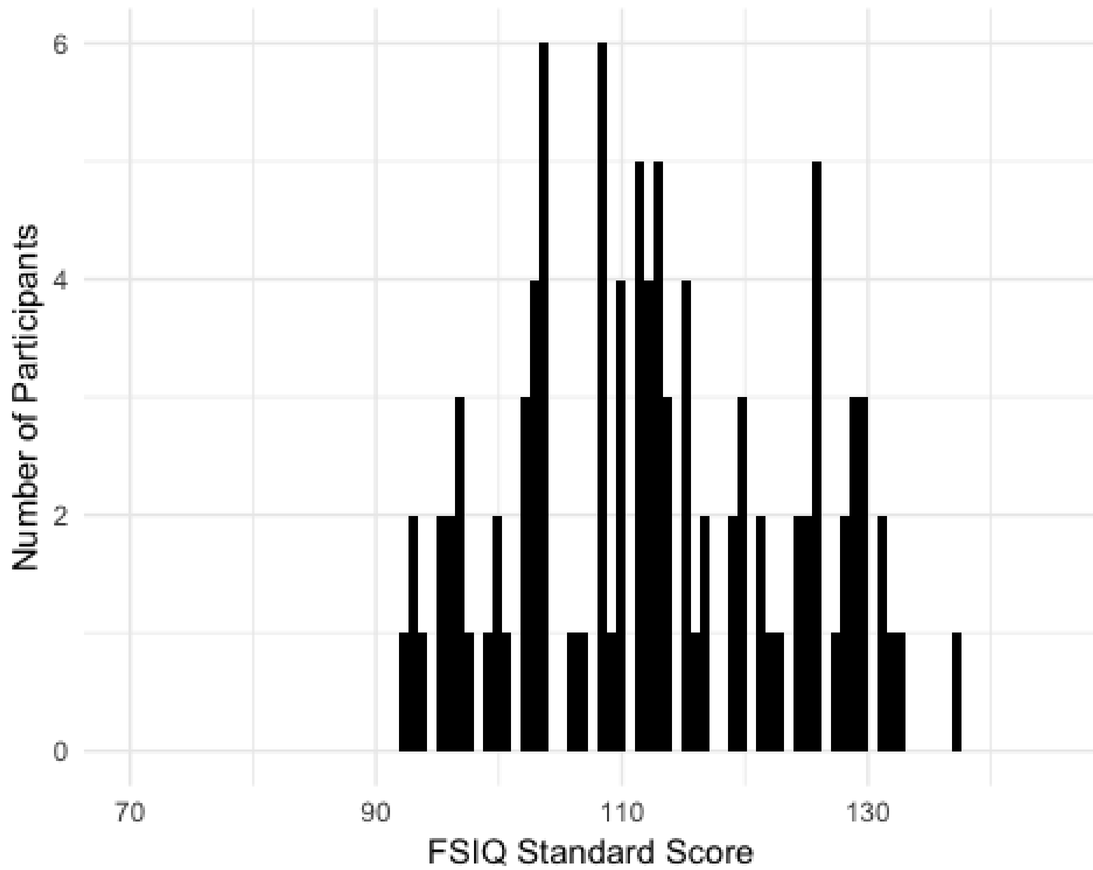


Figure 6. Histogram of FSIQ Scores

```
ggplot(WAISICAR) + aes(x=GAI) + geom_histogram(bins =60, fill = '#000000')  + labs(x = 'GAI Standard Score', y = 'Number of Participants')  + coord_cartesian(xlim=c(70, 145 )) + theme_minimal()

## Warning: Removed 3 rows containing non-finite values (stat_bin).
```

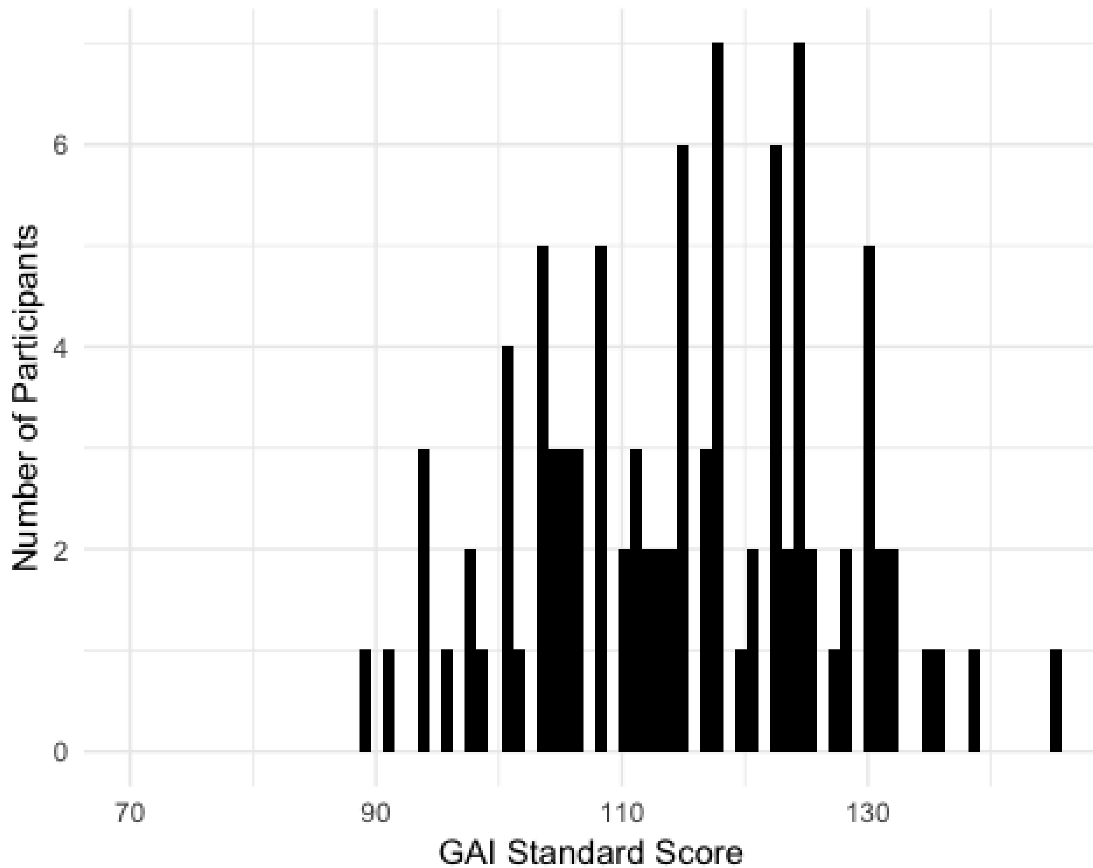


Figure 7. Histogram of GAI Scores

```
ggplot(WAISICAR) + aes(x= TOTAL16) + geom_histogram(bins =16, fill = '#000000')  + labs(x = 'ICAR16 Total Score', y = 'Number of Participants')  + coord_cartesian(xlim=c(0, 16 ), ylim = c(0, 12)) + theme_minimal()
```

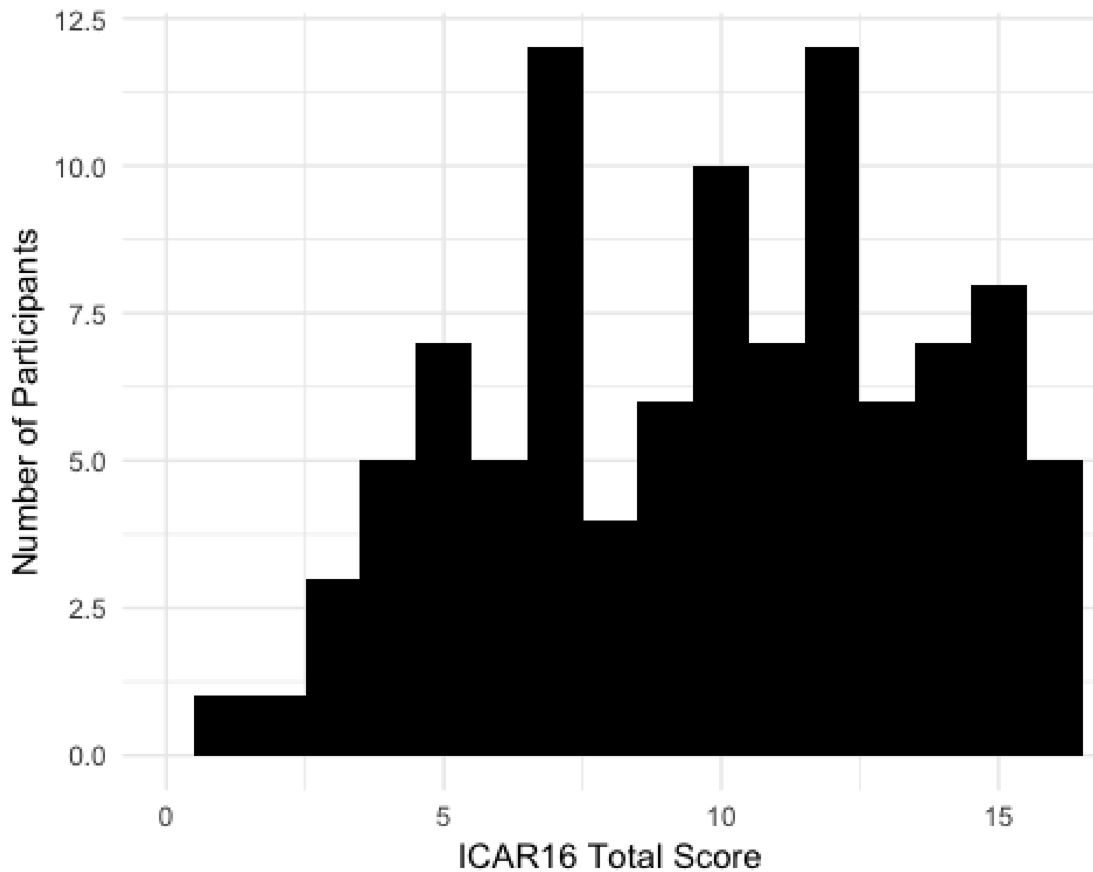


Figure 8. Histogram of ICAR16 Total Scores

```
ggplot(graph_wais_data_gathered) + aes(x=Score) + geom_density()+ facet_wrap(~Test,
ncol=5, nrow=2) + labs(x ='WAIS-IV Subtest Scaled Score', y = 'Participants')  +
coord_cartesian(xlim=c(0, 19)) + theme_minimal()

## Warning: Removed 13 rows containing non-finite values (stat_density).
```

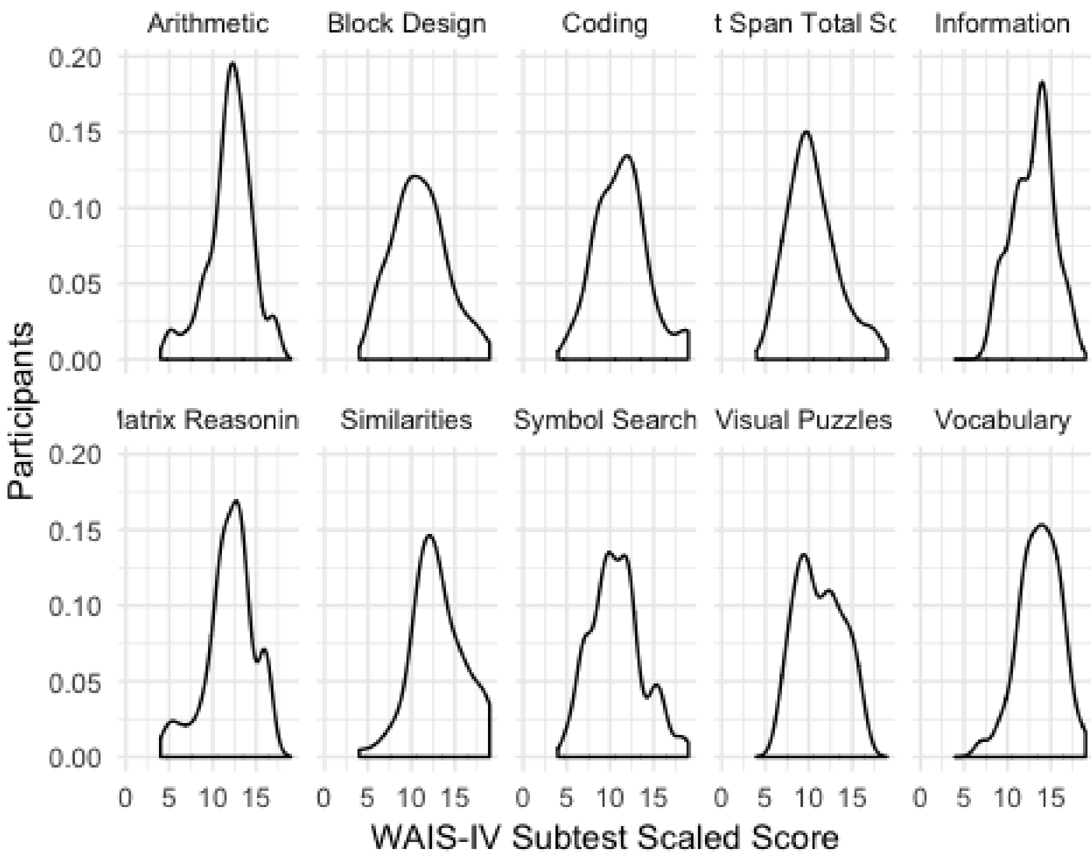


Figure 9. Density Plots of WAIS-IV Subtest Scores

```
ggplot(graph_icar_data) + aes(x=Score) + geom_density() + facet_wrap(~Test, nrow=2) +
labs(x = 'ICAR Subtest Score', y = 'Participants')  + coord_cartesian(xlim=c(0, 4)) +
theme_minimal()
```

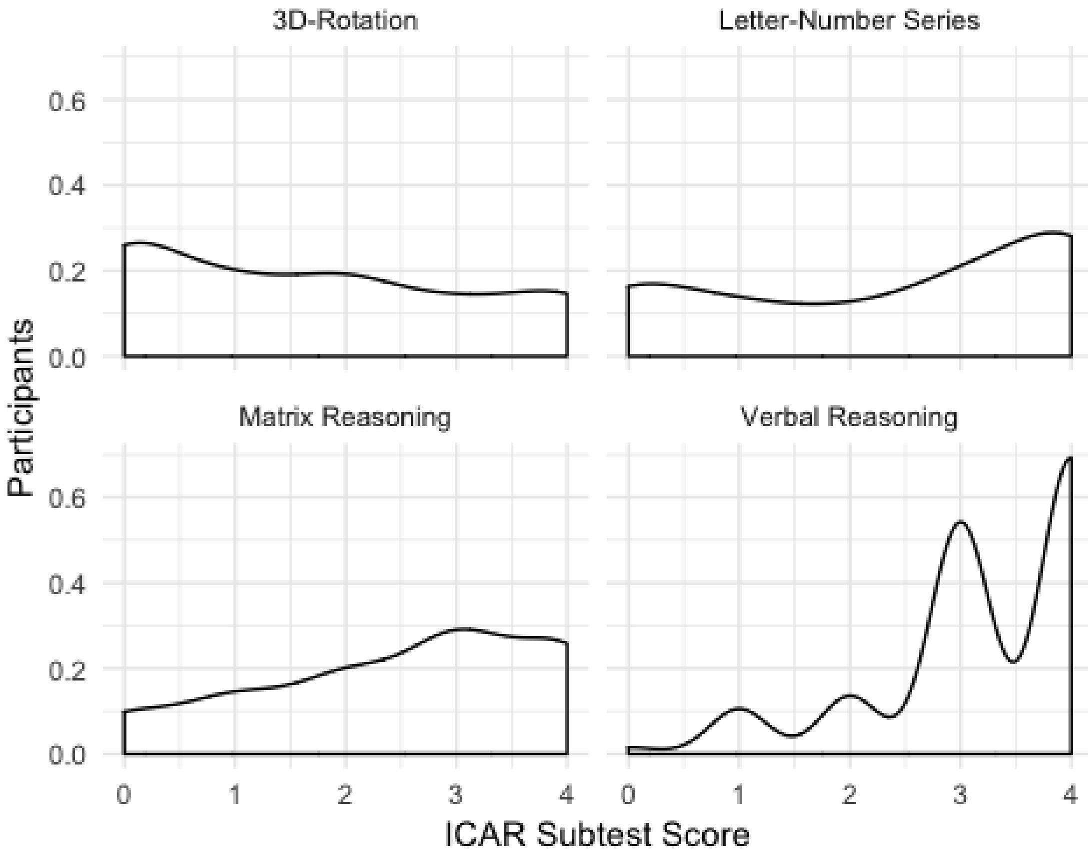


Figure 10. Density Plots of ICAR16 Subtest Scores

```
ggplot(WAISICAR) + aes(x=FSIQ) + geom_density() + labs(x = 'FSIQ Standard Score', y = 'Participants')  + coord_cartesian(xlim=c(70, 145 )) + theme_minimal()

## Warning: Removed 1 rows containing non-finite values (stat_density).
```

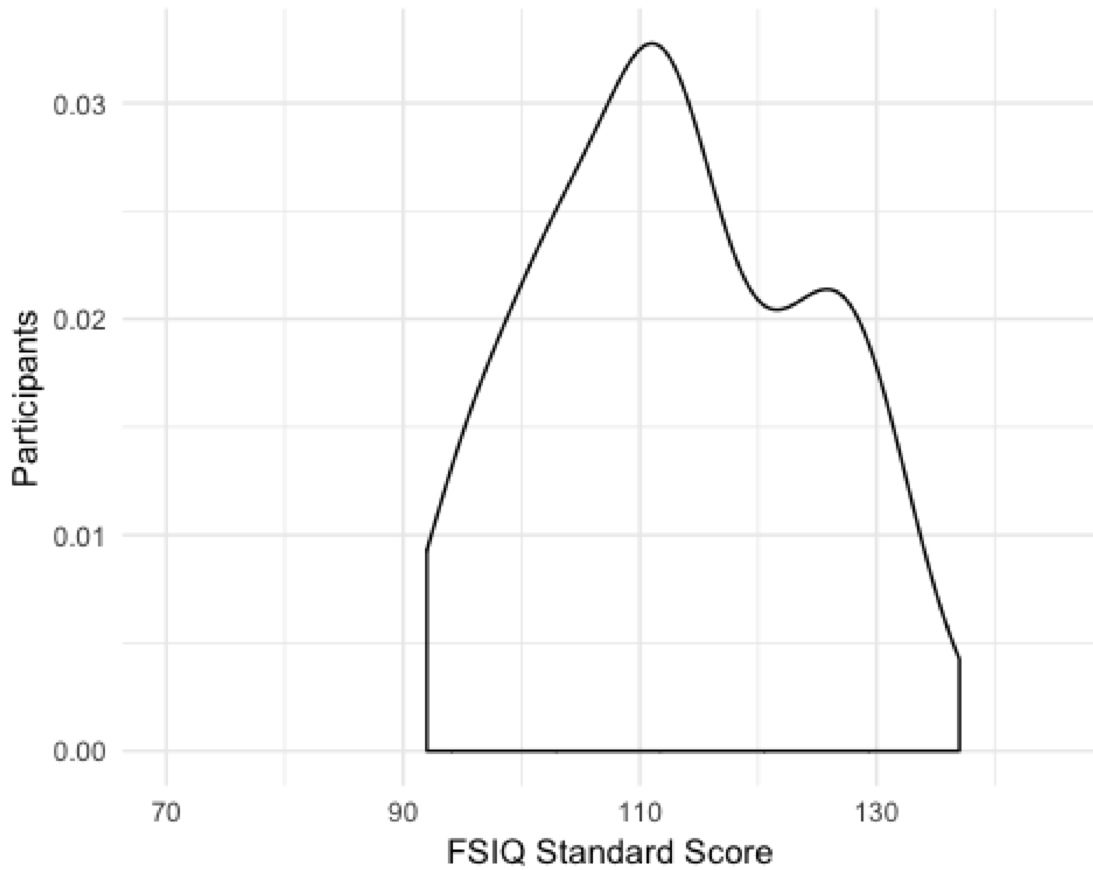


Figure 11. Density Plot of FSIQ Scores

```
ggplot(WAISICAR) + aes(x=GAI) + geom_density()  + labs(x = 'GAI Standard Score', y = 'Participants')  + coord_cartesian(xlim=c(70, 145 )) + theme_minimal()

## Warning: Removed 3 rows containing non-finite values (stat_density).
```

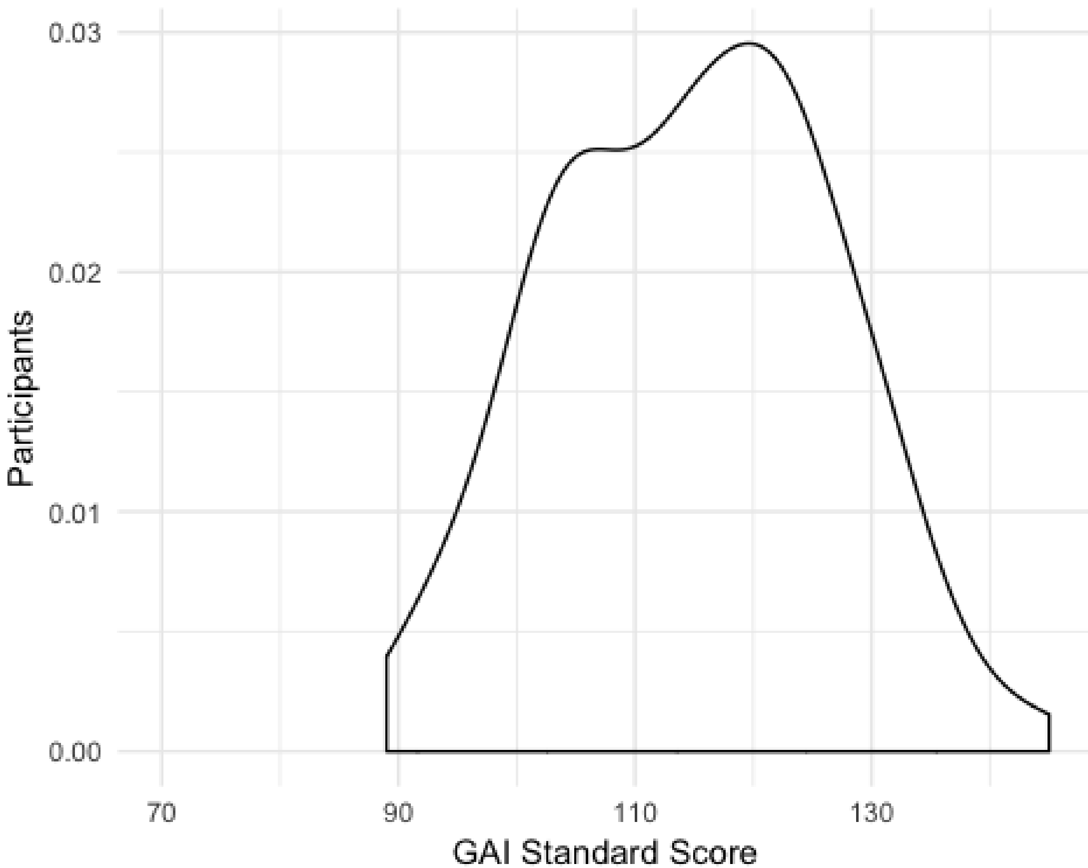


Figure 12. GAI Density Plot

```
ggplot(WAISICAR) + aes(x= TOTAL16) + geom_density()  + labs(x = 'ICAR16 Total Score', y = 'Participants')  + coord_cartesian(xlim=c(0, 16 )) + theme_minimal()
```

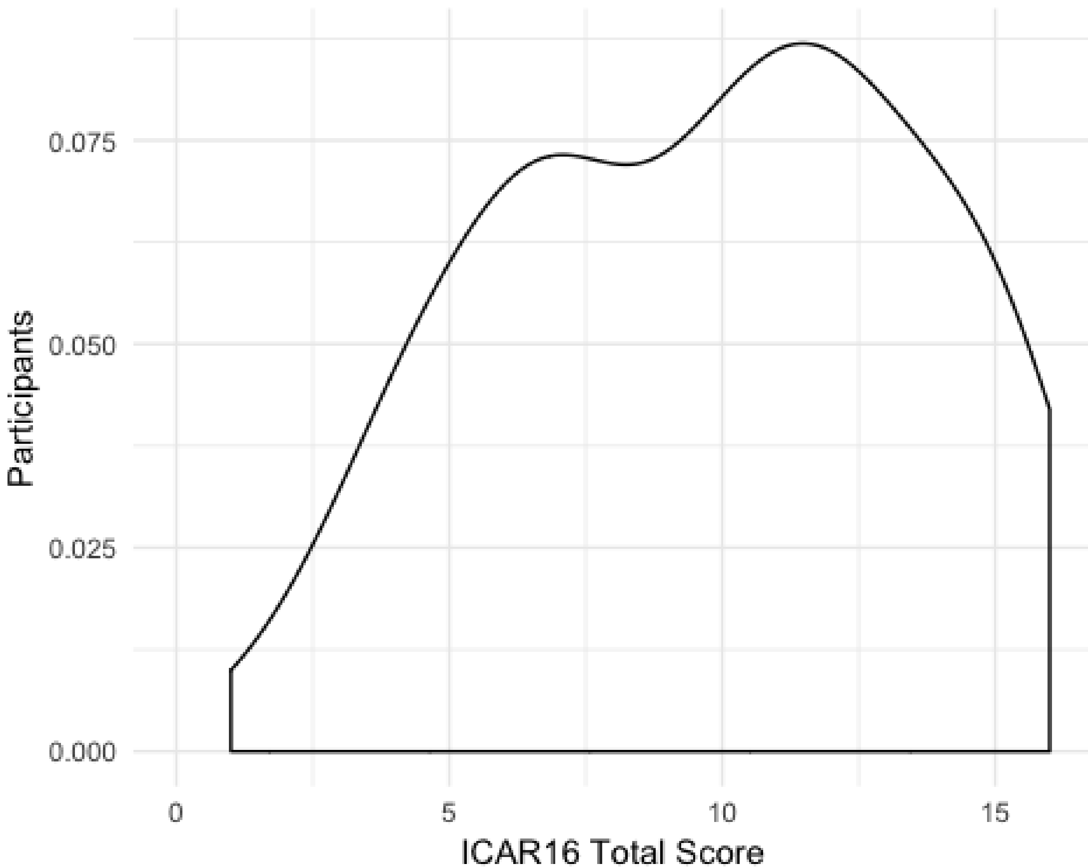


Figure 13. Density Plot ICAR16 Total Score

Check the variances of the overall and composite ability estimates.

```
library(EnvStats)
library(dplyr)
library(tidyr)
library(broom)
library(tidyverse)
get_var<- function(dat) {
  tidy(varTest(dat$Score, sigma.squared = 225)) %>%
     select(estimate, statistic, p.value)
  }

get_var<- function(dat) {
  tidy(varTest(dat$Score, sigma.squared = 225)) %>%
     select(estimate, statistic, p.value)
  }

composites<- WAISICAR %>% dplyr::select(GAI, FSIQ) %>%
  mutate(id = row.names(.)) %>%
  gather(Composite, Score, -id) %>%
```

```
  group_by(Composite)

composites<-na.omit(composites)

styled_comp<-composites %>%
  do(get_var(.))

colnames(styled_comp)<-c("Composite", "Sample Variance", "Chi-Square Difference Test
Statistic", "P-Value")

styled_comp<-styled_comp %>%mutate("Population Variance"= 225)
styled_comp<-styled_comp[c(1, 2, 5, 3, 4)]

styled_comp$`Sample Variance`<-round(styled_comp$`Sample Variance`, 2)
styled_comp$`Chi-Square Difference Test Statistic`<-round(styled_comp$`Chi-Square
Difference Test Statistic`,2)
styled_comp$`P-Value`<-round(styled_comp$`P-Value`,2)

styled_comp<-as.data.frame(styled_comp)

papaja::apa_table(styled_comp, caption="WAIS-IV Composite Scores Sample vs. Population
Variance")
```

Table 15. WAIS-IV Composite Scores Sample vs. Population Variance

Composite	Sample Variance	Population Variance	$\Delta\chi_2$	P-Value
FSIQ	126.63	225.00	54.59	.00
GAI	137.83	225.00	58.19	.00

Check the variances of the WAIS-IV Subtests.

```
library(broom)
library(EnvStats)

get_var<- function(dat) {
  tidy(varTest(dat$Score, sigma.squared = 9)) %>%
    select(estimate, statistic, p.value)
}

standardized_subtests<-na.omit(standardized_subtests)

standardized_subtests_helper<-standardized_subtests %>%
  mutate(id = row.names(.)) %>%
  gather(Subtest, Score, -id) %>%
  group_by(Subtest) %>%
```

```
  do(get_var(.))

colnames(standardized_subtests_helper)<-c("Subtest", "Sample Variance", "Chi-Square
Difference Test Statistic", "P-Value")

standardized_subtests_helper<-standardized_subtests_helper %>%mutate("Population
Variance"= 9)
standardized_subtests_helper<-standardized_subtests_helper[c(1, 2, 5, 3, 4)]

standardized_subtests_helper$`Sample Variance`<-
round(standardized_subtests_helper$`Sample Variance`, 2)

standardized_subtests_helper$`Chi-Square Difference Test Statistic`<-
round(standardized_subtests_helper$`Chi-Square Difference Test Statistic`, 2)

standardized_subtests_helper$`P-Value`<-round(standardized_subtests_helper$`P-Value`, 2)

  papaja::apa_table(standardized_subtests_helper, caption="WAIS-IV Scaled Scores Sample vs.
Population Variance")
```

Table 16. WAIS-IV Scaled Scores Sample vs. Population Variance

Subtest	Sample Variance	Population Variance	$\Delta\chi_2$	p-value
Arithmetic	6.07	9.00	62.68	.01
Block Design	9.63	9.00	99.55	.60
Coding	8.80	9.00	9.94	.92
Digit Span Backward	7.75	9.00	8.09	.34
Digit Span Forward	1.30	9.00	106.44	.32
Digit Span Sequencing	8.85	9.00	91.42	.95
Digit Span Total Score	8.14	9.00	84.16	.53
Information	5.39	9.00	55.67	.00
Matrix Reasoning	7.80	9.00	8.65	.37
Similarities	9.05	9.00	93.54	.93
Symbol Search	8.92	9.00	92.21	.99
Visual Puzzles	6.87	9.00	7.98	.09
Vocabulary	5.69	9.00	58.76	.00

There is significantly less variance in the collected sample for the Information and Vocabulary subtests, which comprise the verbal comprehension broad ability composite (Gc). Thus, a correction for restriction of range should be applied to correlations with the Gc estimate.

Check the variances of the ICAR against a subset of the norming sample with similar demographic characteristics.

```
library(tidyr)
library(dplyr)
library(tidyverse)

completecases<-read_csv("completecases.csv")

## Create subscores from the norming sample
ICARcomplete<- completecases %>%
  mutate(VR16= VR.04 + VR.16 +VR.19 +VR.17, LN16= LN.07 + LN.33 + LN.34 +LN.58,
MX16= MR.45 + MR.46 + MR.47 + MR.55, R3D16= R3D.03 + R3D.04 + R3D.06 +R3D.08,
TOTAL16= VR16 + MX16 + LN16 + R3D16)

## Select cases with similar demographics to collected sample
ICARCOMPLETEMATCH<- ICARcomplete %>%
  filter(age>=18 & age<=45 & education != "less12yrs" & education != "GradOrProDegree" &
status=="student") %>%
  dplyr::select("id", "gender", "age", c(15:39), -c(32:34)) %>%
  #create a new variable to indicate which sample the data comes from
  mutate("SAMPLE_MATCH"='ICARNORM')

## Make sure the column names match exactly
colnames(ICARCOMPLETEMATCH)<-c("ID", "FEMALE", "AGE", "ACT",  "VR_04"  ,
"VR_16" ,   "VR_17",  "VR_19"  , "LN_07" ,  "LN_33", "LN_34", "LN_58", "MR_45", "MR_46",
"MR_47", "MR_55", "R3D_03", "R3D_04", "R3D_06", "R3D_08",  "VR16", "LN16"  ,  "MX16"
,   "R3D16",   "TOTAL16", "SAMPLE_MATCH" )

# Change all the -999 to missing
ICARCOMPLETEMATCH[ICARCOMPLETEMATCH==-999]<- NA

# Subset the collected sample to match variables in the norming subsample
WIMATCH<- WAISICAR %>%
  dplyr::select("ID", "AGE", "FEMALE", "ACT",  "VR_04"  , "VR_16" ,   "VR_17", "VR_19"  ,
"LN_07" ,  "LN_33", "LN_34", "LN_58",  "MR_45", "MR_46", "MR_47", "MR_55", "R3D_03",
"R3D_04", "R3D_06", "R3D_08",   "LN16"  ,  "MX16", "VR16" ,   "R3D16", "TOTAL16" )
%>%
  mutate("SAMPLE_MATCH" = "COLLECTED")
```

```
# Combine the two samples in a new sample
Matchsample<-rbind(WIMATCH, ICARCOMPLETEMATCH)
```

Compare the variances of the two samples.

```
library(car)
Matchsample$SAMPLE_MATCH<-as.factor(Matchsample$SAMPLE_MATCH)

#Total score
TS<-tidy(leveneTest(Matchsample$TOTAL16 ~ Matchsample$SAMPLE_MATCH)) %>%
mutate(variable="TOTAL16") %>% filter(df==1) %>% dplyr::select(c("variable", "statistic",
"p.value"))

#LN16
LN<-tidy(leveneTest(Matchsample$LN16 ~ Matchsample$SAMPLE_MATCH)) %>%
mutate(variable="LN16") %>% filter(df==1) %>% dplyr::select(c("variable", "statistic",
"p.value"))

#VR16
VR<- tidy(leveneTest(Matchsample$VR16 ~ Matchsample$SAMPLE_MATCH)) %>%
mutate(variable="VR16") %>% filter(df==1) %>% dplyr::select(c("variable", "statistic",
"p.value"))

#MX16
MR<-tidy(leveneTest(Matchsample$MX16 ~ Matchsample$SAMPLE_MATCH)) %>%
mutate(variable="MX16") %>% filter(df==1) %>% dplyr::select(c("variable", "statistic",
"p.value"))

#R3D16
R3D<-tidy(leveneTest(Matchsample$R3D16 ~ Matchsample$SAMPLE_MATCH)) %>%
mutate(variable="R3D16") %>% filter(df==1) %>% dplyr::select(c("variable", "statistic",
"p.value"))

Levene_tests<-bind_rows( LN, VR, MR, R3D,TS)
colnames(Levene_tests)<-c("ICAR Subtest", "Levene Test Statistic", "P-Value")

Levene_tests$`Levene Test Statistic`<-round(Levene_tests$`Levene Test Statistic`, 2)
Levene_tests$`P-Value`<-round(Levene_tests$`P-Value`,2)

  papaja::apa_table( Levene_tests, caption="ICAR Subtest Scores Sample vs. Population
Variance")
```

Table 17. ICAR Subtest Scores Sample vs. Population Variance

ICAR Subtest	Levene Test Statistic	p-value
LN16	1.81	.18
VR16	8.94	.00
MX16	.38	.54
R3D16	7.62	.01
TOTAL16	.47	.49

There is significantly less variance in the R3D and VR tasks in the collected sample when compared to the demographic-matched subset from the norming sample.

Appendix B: Correlation Tables Individual WAIS-IV Subtests and ICAR16

Table 18. Uncorrected correlations with confidence intervals, WAIS-IV raw subtest scores and ICAR16

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. AR	16.41	3.08													
2. MR	21.14	3.67	.29** [.10, .47]												
3. BD	48.21	10.82	.30** [.11, .47]	.29** [.10, .46]											
4. VP	17.77	3.97	.27** [.08, .45]	.38** [.19, .54]	.61** [.47, .72]										
5. DS	29.06	5.31	.45** [.27, .59]	.23* [.04, .41]	.17 [-.03, .35]	.27** [.08, .44]									
6. VC	44.03	7.14	.31** [.11, .48]	.15 [-.05, .34]	.12 [-.08, .31]	.15 [-.05, .34]	.36** [.18, .52]								
7. SI	28.96	4.26	.36** [.18, .52]	.27** [.07, .44]	.16 [-.04, .35]	.25* [.05, .42]	.27** [.08, .45]	.35** [.16, .51]							
8. IN	17.71	3.41	.44** [.26, .59]	.34** [.15, .51]	.24* [.04, .42]	.27** [.08, .45]	.22* [.02, .41]	.58** [.43, .70]	.30** [.11, .47]						
9. SS	35.33	8.41	.17 [-.03, .36]	.11 [-.09, .30]	.29** [.10, .46]	.20* [.00, .38]	.19 [-.01, .37]	.08 [-.12, .28]	.14 [-.06, .33]	-.04 [-.24, .16]					
10. CD	79.39	15.35	.20* [.00, .39]	.27** [.08, .45]	.18 [-.02, .36]	.23* [.04, .41]	.16 [-.04, .35]	.14 [-.06, .33]	.27** [.07, .44]	.24* [.04, .42]	.50** [.34, .64]				
11. LN16	2.37	1.56	.49** [.32, .63]	.44** [.26, .58]	.36** [.17, .52]	.30** [.11, .47]	.28** [.09, .45]	.23* [.03, .41]	.18 [-.02, .36]	.42** [.24, .57]	.23* [.03, .41]	.33** [.14, .50]			
12. MX16	2.49	1.30	.32** [.13, .49]	.29** [.10, .46]	.34** [.15, .51]	.45** [.27, .59]	.27** [.07, .44]	.20* [.00, .38]	.16 [-.04, .34]	.26* [.06, .44]	.16 [-.04, .34]	.23* [.03, .41]	.41** [.24, .57]		
13. VR16	3.20	0.95	.25* [.05, .42]	.19 [-.01, .38]	.42** [.24, .57]	.41** [.23, .57]	.36** [.18, .52]	.17 [-.03, .36]	.33** [.14, .50]	.27** [.07, .44]	.11 [-.09, .30]	.05 [-.15, .24]	.37** [.19, .53]	.34** [.15, .50]	
14. R3D16	1.67	1.47	.24* [.05, .42]	.12 [-.08, .31]	.31** [.12, .48]	.32** [.13, .49]	.22* [.02, .40]	.32** [.13, .49]	.19 [-.01, .37]	.21* [.01, .40]	.02 [-.18, .21]	.10 [-.10, .29]	.43** [.25, .58]	.40** [.21, .55]	.18 [-.02, .37]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

Table 19. Uncorrected correlations with confidence intervals, WAIS-IV scaled subtest scores and ICAR16

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. AR	12.01	2.53													
2. MR	11.99	2.78	.36** [.18, .53]												
3. BD	10.96	3.09	.31** [.12, .48]	.33** [.14, .50]											
4. VP	11.26	2.61	.29** [.10, .46]	.42** [.24, .57]	.61** [.47, .72]										
5. DS	10.42	2.85	.43** [.26, .58]	.21* [.01, .39]	.20 [-.00, .38]	.31** [.12, .48]									
6. VC	13.71	2.39	.30** [.10, .47]	.12 [-.08, .31]	.09 [-.11, .29]	.16 [-.04, .34]	.31** [.12, .48]								
7. SI	13.04	3.01	.38** [.19, .53]	.21* [.01, .39]	.19 [-.01, .37]	.25* [.06, .43]	.22* [.02, .40]	.34** [.15, .50]							
8. IN	12.97	2.32	.44** [.26, .59]	.30** [.11, .48]	.27** [.07, .44]	.30** [.10, .47]	.21* [.01, .39]	.56** [.41, .69]	.23* [.03, .41]						
9. SS	10.78	2.99	.11 [-.09, .30]	.06 [-.14, .25]	.30** [.11, .47]	.20 [-.00, .38]	.19 [-.01, .38]	.05 [-.15, .25]	.14 [-.06, .33]	-.03 [-.23, .17]					
10. CD	11.38	3.00	.19 [-.00, .38]	.25* [.06, .43]	.18 [-.02, .37]	.26* [.06, .44]	.16 [-.05, .34]	.08 [-.12, .28]	.22* [.03, .41]	.18 [-.02, .37]	.46** [.29, .61]				
11. LN16	2.37	1.56	.49** [.32, .62]	.44** [.26, .59]	.35** [.17, .52]	.31** [.12, .48]	.26** [.07, .44]	.24* [.05, .42]	.18 [-.02, .37]	.40** [.21, .55]	.21* [.01, .39]	.33** [.14, .50]			
12. MX16	2.49	1.30	.33** [.14, .49]	.29** [.09, .46]	.31** [.12, .48]	.45** [.27, .59]	.27** [.07, .44]	.21* [.01, .39]	.21* [.02, .40]	.26* [.06, .44]	.12 [-.08, .31]	.22* [.03, .41]	.41** [.24, .57]		
13. VR16	3.20	0.95	.24* [.04, .42]	.18 [-.01, .37]	.38** [.20, .54]	.40** [.22, .55]	.35** [.16, .51]	.22* [.02, .40]	.32** [.13, .49]	.28** [.09, .46]	.14 [-.06, .33]	.04 [-.17, .23]	.37** [.19, .53]	.34** [.15, .50]	
14. R3D16	1.67	1.47	.26** [.06, .44]	.13 [-.07, .32]	.30** [.11, .47]	.32** [.14, .49]	.21* [.01, .39]	.26* [.06, .43]	.18 [-.02, .37]	.22* [.03, .41]	.05 [-.15, .25]	.05 [-.15, .25]	.43** [.25, .58]	.40** [.21, .55]	.18 [-.02, .37]

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

Table 20. Range corrected correlations with confidence intervals, WAIS-IV scaled subtest scores and ICAR16

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. AR													
2. MR	.42** [.24, .57]												
3. BD	.36** [.18, .52]	.32** [.13, .49]											
4. VP	.34** [.15, .5]	.44** [.27, .59]	.60** [.46, .71]										
5. DS	.50** [.33, .63]	.22* [.02, .4]	.19 [-.01, .37]	.33** [.14, .49]									
6. VC	.35** [.16, .51]	.13 [-.07, .32]	.09 [-.11, .28]	.20 [0, .38]	.32** [.13, .49]								
7. SI	.43** [.26, .58]	.23* [.03, .41]	.18 [-.01, .37]	.25* [.06, .43]	.23* [.03, .41]	.34** [.15, .5]							
8. IN	.50** [.34, .64]	.38** [.2, .54]	.26*** [.07, .44]	.37** [.19, .53]	.22* [.02, .4]	.66** [.54, .76]	.30** [.11, .47]						
9. SS	.13 [-.07, .32]	.06 [-.14, .26]	.29** [.1, .46]	.20 [0, .38]	.20 [0, .38]	.05 [-.14, .25]	.14 [-.06, .33]	-.04 [-.23, .16]					
10. CD	.23* [.03, .41]	.25* [.06, .43]	.18 [-.02, .36]	.26* [.06, .43]	.15 [-.04, .34]	.08 [-.12, .28]	.22* [.03, .4]	.18 [-.02, .37]	.46** [.29, .61]				
11. LN16	.55** [.4, .67]	.40** [.21, .55]	.34** [.16, .51]	.28** [.09, .45]	.27** [.08, .45]	.22* [.02, .4]	.16 [-.04, .35]	.49** [.32, .63]	.18 [-.02, .37]	.33** [.14, .5]			
12. MX16	.38** [.2, .54]	.3188 [.12, .48]	.30** [.11, .47]	.43** [.25, .58]	.28** [.09, .45]	.20* [.01, .38]	.20* [.01, .39]	.33** [.14, .49]	.11 [-.08, .31]	.22* [.03, .4]	.37** [.19, .53]		
13. VR16	.28** [.09, .45]	.20 [0, .38]	.37** [.19, .53]	.45** [.27, .59]	.37** [.18, .53]	.27** [.08, .45]	.32** [.13, .48]	.36** [.17, .52]	.14 [-.06, .33]	.04 [-.16, .23]	.33** [.14, .5]	.32** [.14, .49]	
14. R3D16	.30** [.11, .47]	.14 [-.06, .32]	.29** [.1, .46]	.27** [.08, .45]	.22* [.02, .4]	.21* [.02, .39]	.15 [-.05, .34]	.29** [.09, .46]	.04 [-.16, .24]	.05 [-.15, .25]	.38** [.2, .54]	.38** [.2, .54]	.15 [-.05, .34]

Note. Restriction of range was corrected by using formula (1) from Bryant and Gohale (1972) and Alexander (1990). manual (David Wechsler, 2008) and the initial ICAR validation study (Condon & Revelle, 2014). Values in square brackets indicate the 95% confidence interval for each range corrected correlation. * indicates $p < .05$. ** indicates $p < .01$.

Table 21. Range-and-Reliability Corrected Correlations and Confidence Intervals

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. AR	.88													
2. MR	.47**	.9												
	[.31, .61]													
3. BD	.41**	.36**	.87											
	[.24, .57]	[.18, .52]												
4. VP	.38**	.50**	0.68**	.89										
	[.2, .54]	[.33, .63]	[.56, .78]											
5. DS	.55**	.24*	.21*	.36**	.93									
	[.39, .67]	[.04, .42]	[.02, .39]	[.18, .52]										
6. VC	.38**	.14	.10	.21*	.34**	.94								
	[.2, .54]	[-.06, .33]	[-.1, .29]	[.02, .39]	[.16, .51]									
7. SI	.50**	.26*	.21*	.28**	.26*	.38**	.87							
	[.33, .63]	[.06, .43]	[.01, .39]	[.09, .46]	[.06, .43]	[.19, .53]								
8. IN	.55**	.42**	.29**	.41**	.23*	.71**	.33**	.93						
	[.4, .68]	[.24, .57]	[.1, .46]	[.23, .56]	[.04, .41]	[.59, .79]	[.14, .49]							
9. SS	.16	.07	.35**	.23*	.23*	.06	.17**	-.04	.81					
	[-.04, .34]	[-.13, .27]	[.16, .51]	[.04, .41]	[.04, .41]	[-.14, .26]	[-.03, .35]	[-.24, .16]						
10. CD	.26**	.29**	.21*	.30**	.17	.09	.26*	.20*	.55**	.86				
	[.07, .44]	[.1, .46]	[.01, .39]	[.1, .47]	[-.02, .36]	[-.11, .28]	[.07, .43]	[.01, .39]	[.4, .68]					
11. LN16	.71**	.51**	.45**	.36**	.35**	.27**	.21*	.61**	.25*	.43**	.68			
	[.6, .8]	[.34, .64]	[.28, .59]	[.17, .52]	[.16, .51]	[.08, .44]	[.01, .39]	[.47, .72]	[.05, .42]	[.26, .58]				
12. MX16	.56**	.45**	.45**	.63**	.40**	.29**	.30**	.47**	.18	.33**	.63**	.52		
	[.41, .69]	[.28, .6]	[.27, .59]	[.5, .74]	[.22, .56]	[.1, .46]	[.11, .47]	[.3, .61]	[-.02, .36]	[.15, .5]	[.49, .73]			
13. VR16	.39**	.27**	.52**	.62**	.49**	.36**	.44**	.48**	.20*	.05	.52**	.59**	.59	
	[.21, .55]	[.08, .45]	[.36, .65]	[.48, .73]	[.33, .63]	[.18, .52]	[.27, .59]	[.31, .62]	[0, .38]	[-.15, .24]	[.36, .65]	[.44, .7]		
14. R3D16	.38**	.17	.37**	.34**	.26**	.26*	.19	.34**	.05	.06	.54**	.61**	.23*	.74
	[.11, .47]	[-.03, .35]	[.18, .53]	[.15, .5]	[.07, .44]	[.06, .43]	[-.01, .37]	[.16, .51]	[-.15, .25]	[-.14, .26]	[.39, .67]	[.47, .72]	[.03, .41]	

Note. Formula (2) from Murphy and Davidshofer** (1988) applied to the range-corrected correlations in Table 9 using the reliabilities published in the WAIS-IV technical manual (David Wechsler, 2008) and the initial ICAR validation study (Condon & Revelle, 2014). * indicates $p < .05$. ** indicates $p < .01$.

* indicates $p < .05$. ** indicates $p < .01$

Appendix C: Informed Consent Documents

The University of Texas at Austin
Psychological and Educational Assessment Center (PEAC)

Introduction
The purpose of this form is to provide you information that may affect your decision as to whether or not to participate in this research study. The person performing the research will answer any of your questions. Read the information below and ask any questions you might have before deciding whether or not to take part. If you decide to be involved in this study, this form will be used to record your consent.

Purpose of the Study
You have been asked to participate in a research study about the tools used to assess cognitive and academic abilities. The purpose of this study is to better understand the properties of the tests used to evaluate cognitive and academic abilities.

What will you be asked to do?
If you agree to participate in this study, you will be asked for your consent to have your assessment data analyzed for research purposes. Assessment information is protected under the Family Educational Rights and Privacy Act (FERPA), which protects the privacy of student educational records. Findings from this study will be included in research presentations or papers for publication. This study will take not take any of your time.

What are the risks involved in this study?
There are no foreseeable risks to participating in this study.

What are the possible benefits of this study?
You will receive no direct benefit from participating in this study; however, the findings from this study will advance our ability to assess cognitive and academic abilities.

Do you have to participate?
No, your participation is voluntary. You may decide not to participate at all or, if you consent to allow your data to be used in the study, you may withdraw consent at any time. Withdrawal or refusing to participate will not affect your relationship with The University of Texas at Austin in any way.

Will there be any compensation?
You will not receive any type of payment participating in this study.

How will your privacy and confidentiality be protected if you participate in this research study?
Your privacy and the confidentiality of your data will be protected by storage on Qualtrics and UT Box. Any identifying information will be removed and replaced with a unique identifier. Qualtrics and UT Box have been approved by the University's Information Security Office for use with Confidential (formerly known as Category I) university data, including HIPAA data. Only researchers involved in the study will have access to this data. If it becomes necessary for the Institutional Review Board to review the study records, information that can be linked to you

will be protected to the extent permitted by law. Your research records will not be released without your consent unless required by law or a court order. The data resulting from your participation may be made available to other researchers in the future for research purposes not detailed within this consent form. In these cases, the data will contain no identifying information that could associate it with you, or with your participation in any study.

Whom to contact with questions about the study?
Prior, during, or after your participation you can contact the researcher Dr. Timothy Keith or send an email to tzkeith@austin.utexas.edu for any questions or if you feel that you have been harmed.

Whom to contact with questions concerning your rights as a research participant?
For questions about your rights or any dissatisfaction with any part of this study, you can contact, anonymously if you wish, the Institutional Review Board by phone at (512) 471-8871 or email at orsc@uts.cc.utexas.edu.

Participation
By clicking the button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

- I consent. I agree to allow my de-identified assessment data to be used for research purposes.
- I do not consent. I do not want my de-identified assessment data used for research purposes.

Educational Psychology Subject Pool and Volunteer Participants
Consent for Participation in Research

Title: Investigating the Convergent Validity of the ICAR and WAIS-IV

Introduction

The purpose of this form is to provide you information that may affect your decision as to whether or not to participate in this research study. The person performing the research will answer any of your questions. Read the information below and ask any questions you might have before deciding whether or not to take part. If you decide to be involved in this study, this form will be used to record your consent.

Purpose of the Study

The purpose of this research study is to examine the validity of a public domain measure of cognitive ability. Your participation in the study will contribute to a better understanding of freely accessible cognitive ability research tools. You are free to contact the investigator at the above address and phone number to discuss the study. You must be at least 18 years old to participate.

What will you be asked to do?

If you agree to participate in this study, you will be asked to

- You will complete an intake form that will gather information about your demographic background and educational experiences
- You will complete a multiple-choice test of cognitive ability online.
- You will complete an in-person assessment of cognitive ability with a trained graduate student administrator at the Psychological and Educational Assessment Center in the George I. Sanchez Building on the University of Texas at Austin Campus.

This study will take an estimated 90-120 minutes and will include approximately 80 study participants.

What are the risks involved in this study?

There are no foreseeable risks to participating in this study.

What are the possible benefits of this study?

You will receive no direct benefit from participating in this study; however, you will contribute to research on freely accessible measurement tools that advance research in health care and the social sciences.

Do you have to participate?

No, your participation is voluntary. You may decide not to participate at all or, if you start the study, you may withdraw at any time. Withdrawal or refusing to participate will not affect your relationship with The University of Texas at Austin (University) in anyway.

If you would like to participate you will receive a copy of this form.

Will there be any compensation?

You will not receive any type of payment participating in this study. Students who choose to participate via the Department of Educational Psychology subject pool may receive course credit for their participation. If they choose not to participate they can complete the alternative options addressed in their class course.

How will your privacy and confidentiality be protected if you participate in this research study?

Your privacy and the confidentiality of your data will be protected by removing identifying information from your data. Your data will be securely stored on UT Box, which meets security requirements for UT Category 1 data.

If it becomes necessary for the Institutional Review Board to review the study records, information that can be linked to you will be protected to the extent permitted by law. Your research records will not be released without your consent unless required by law or a court order. The data resulting from your participation may be made available to other researchers in the future for research purposes not detailed within this consent form. In these cases, the data will contain no identifying information that could associate it with you, or with your participation in any study.

Whom to contact with questions about the study?

Prior, during or after your participation you can contact the researcher Stephanie Young at (612) 819-4550 or send an email to Stephanie.young@utexas.edufor any questions or if you feel that you have been harmed.

Whom to contact with questions concerning your rights as a research participant?

For questions about your rights or any dissatisfaction with any part of this study, you can contact, anonymously if you wish, the Institutional Review Board by phone at (512) 471-8871 or email at orsc@uts.cc.utexas.edu.

Participation

If you agree to participate, sign this form and return it to the researcher.

Signature

You have been informed about this study's purpose, procedures, possible benefits and risks, and you have received a copy of this form. You have been given the opportunity to ask questions before you sign, and you have been told that you can ask other questions at any time. You voluntarily agree to participate in this study. By signing this form, you are not waiving any of your legal rights.

____
Printed Name

__________________________________ __________________
Signature Date

As a representative of this study, I have explained the purpose, procedures, benefits, and the risks involved in this research study.

____
Print Name of Person obtaining consent

_____________________________________ ____________________
Signature of Person obtaining consent Date

Appendix D: Demographic Survey

What is your date of birth? (Please enter MM/DD/YYYY)

How do you identify your gender?
- Male
- Female
- Other

How do you identify your race? If you are more than one race, please select all that apply.
- American Indian or Alaska Native
- Asian
- Black or African American
- Native Hawaiian or Other Pacific Islander
- White
- Other (please specify) ____

Are you of Hispanic, Latino, or Spanish origin?
- Yes
- No

What was the primary language you spoke in the home growing up?
- English
- Spanish
- Another language
- English and Spanish equally
- Another combination of languages equally

What language(s)?

What is the highest level of education achieved by your mother?
- Less than a high school diploma
- High school diploma
- Some college or university experience
- Associate degree
- Bachelor's degree
- Master's degree
- Professional or Doctoral Degree (e.g. M.D., J.D., Ph.D. etc.)
- I don't know

What is the highest level of education achieved by your father?
- Less than a high school diploma
- High school diploma
- Some college or university experience
- Associate degree

- Bachelor's degree
- Master's degree
- Professional or Doctoral Degree (e.g. M.D., J.D., Ph.D. etc.)
- I don't know

How many years have you attended the University of Texas at Austin?
- 1
- 2
- 3
- 4
- 5+

Are you a transfer student?
- Yes
- No

How many years at UT do you expect it will take you to graduate?
- 3
- 4
- 5+

What is your current major?

What is your current college GPA? If you are unsure, please estimate.

What was your final high school GPA? If you are unsure, please estimate.

What was your final class rank in high school? If you are unsure, please estimate. (Please format your answer: rank position/size of graduating class)

What high school did you attend? (If multiple, enter the school from which you earned your diploma)

In what zip code did you attend high school?

Did you take the...
- SAT
- ACT
- Neither

What was your score on the SAT?

What was your score on the ACT?

Have you previously undergone a psychological or educational evaluation?

- Yes
- No

Please describe the results of the evaluation.

Did you receive special education services in a K-12 setting?

- Yes
- No

Did you receive 504 accommodation services in a K-12 setting?

- Yes
- No

What is the primary reason you are seeking assessment services?
Please select

- ADHD
- Dyscalculia
- Dyslexia
- Dysgraphia
- General learning difficulties
- Anxiety
- Other

Please briefly describe the primary reason you are seeking assessment services.

Appendix E: The International Cognitive Ability Resource Sample Test (ICAR16)

What number is one fifth of one fourth of one ninth of 900?

- 2
- 3
- 4
- 5
- 6
- 7

Zach is taller than Matt and Richard is shorter than Zach. Which of the following statements would be most accurate?

- Richard is taller than Matt
- Richard is shorter than Matt
- Richard is as tall as Matt
- It is impossible to tell

Joshua is 12 years old and his sister is three times as old as he. When Joshua is 23 years old, how old will his sister be?

- 35
- 39
- 44
- 47
- 53
- 57

If the day after tomorrow is two days before Thursday, then what day is it today?

- Friday
- Monday
- Wednesday
- Saturday
- Tuesday
- Sunday

In the following alphanumeric series, what letter comes next? K N P S U

- S
- T
- U
- V
- W
- X

In the following alphanumeric series, what letter comes next? V Q M J H

- E
- F

- G
- H
- I
- J

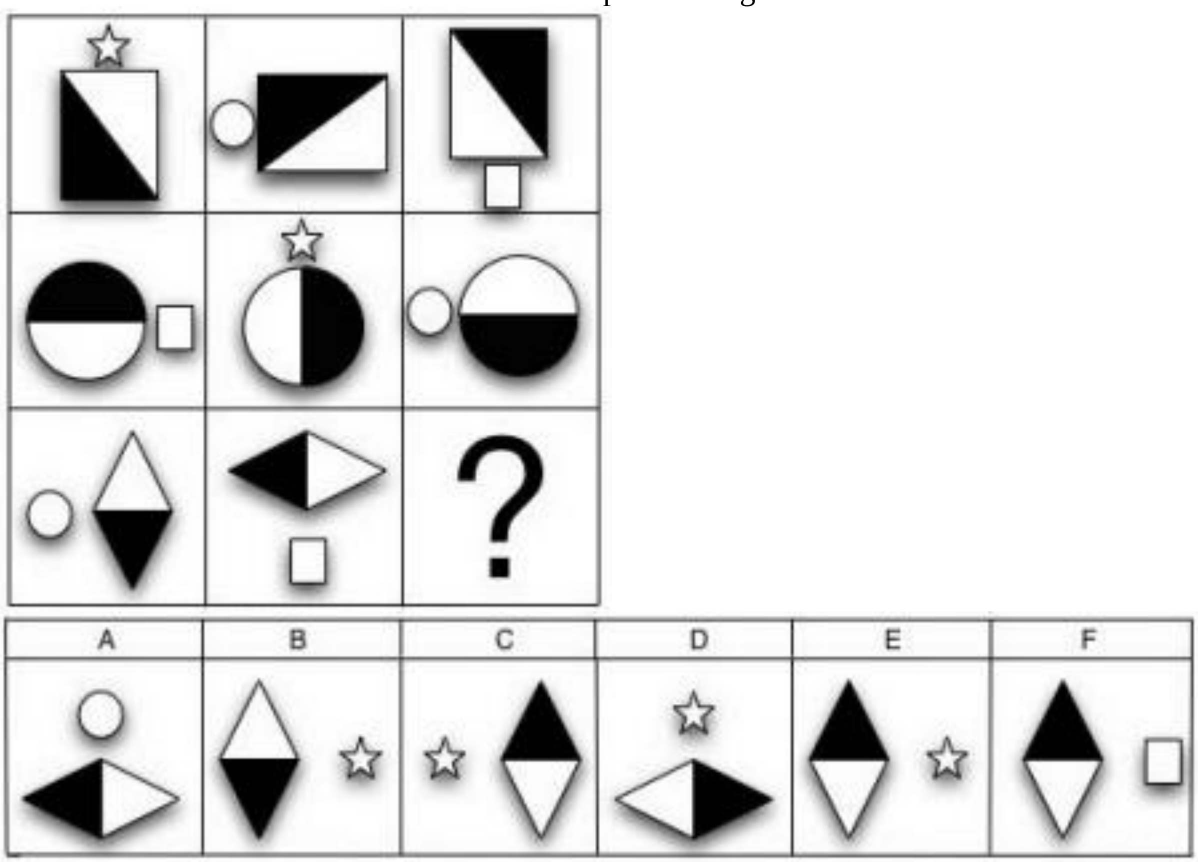
In the following alphanumeric series, what letter comes next? I J L O S

- T
- U
- V
- X
- Y
- Z

In the following alphanumeric series, what letter comes next? Q S N P L

- J
- H
- I
- N
- M
- L

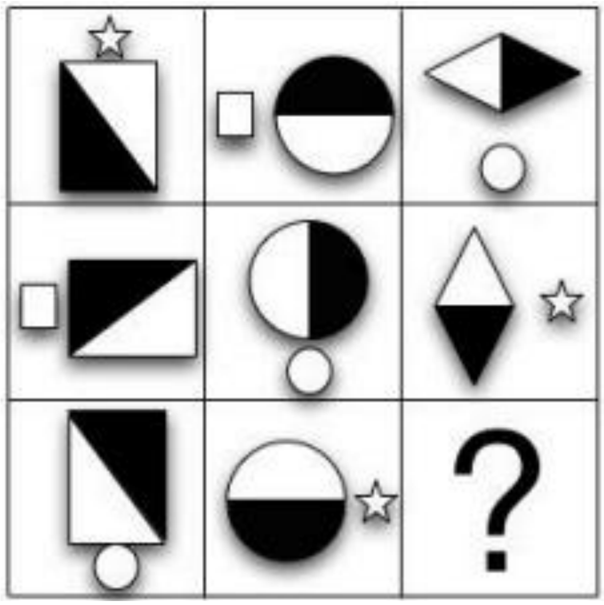
Please indicate which is the best answer to complete the figure below

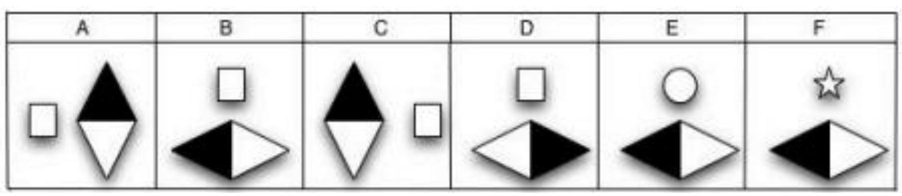


- B

- C
- D
- E
- F

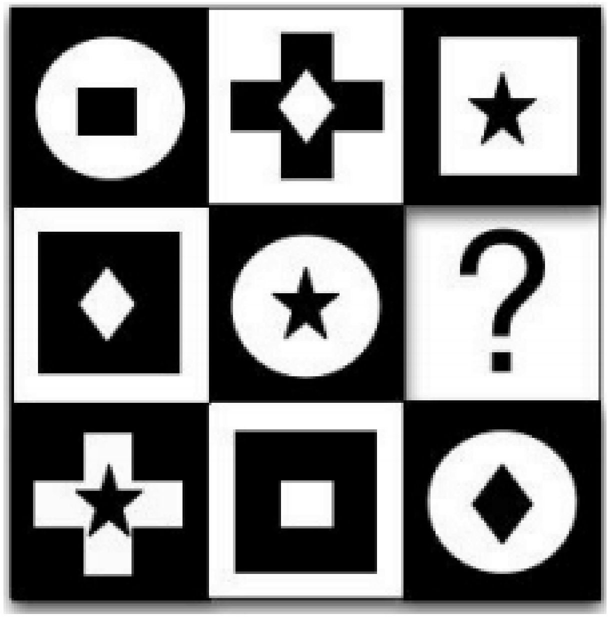
Please indicate which is the best answer to complete the figure below.

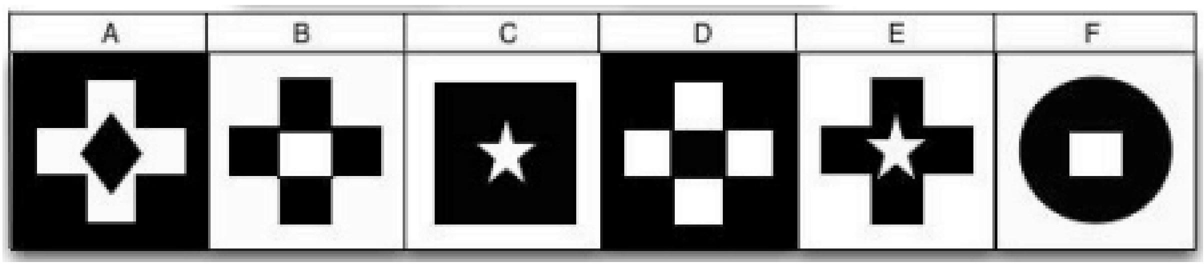




- A
- B
- C
- D
- E
- F

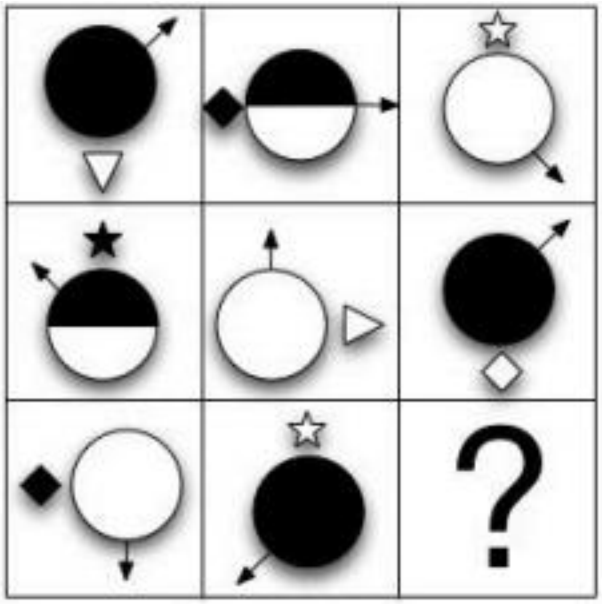
Please indicate which is the best answer to complete the figure below.

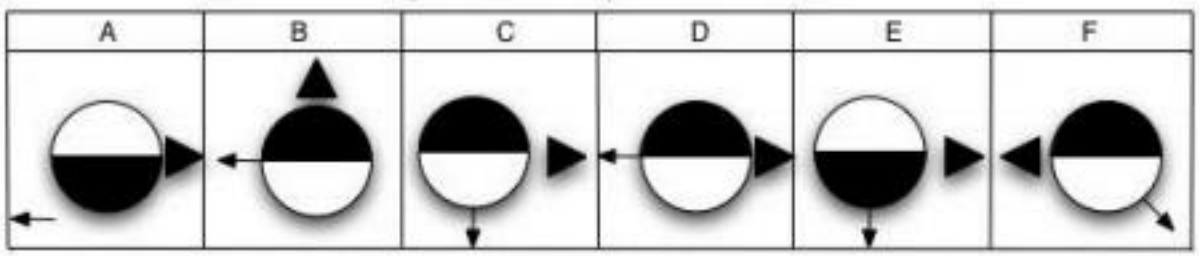




- A
- B
- C
- D
- E
- F

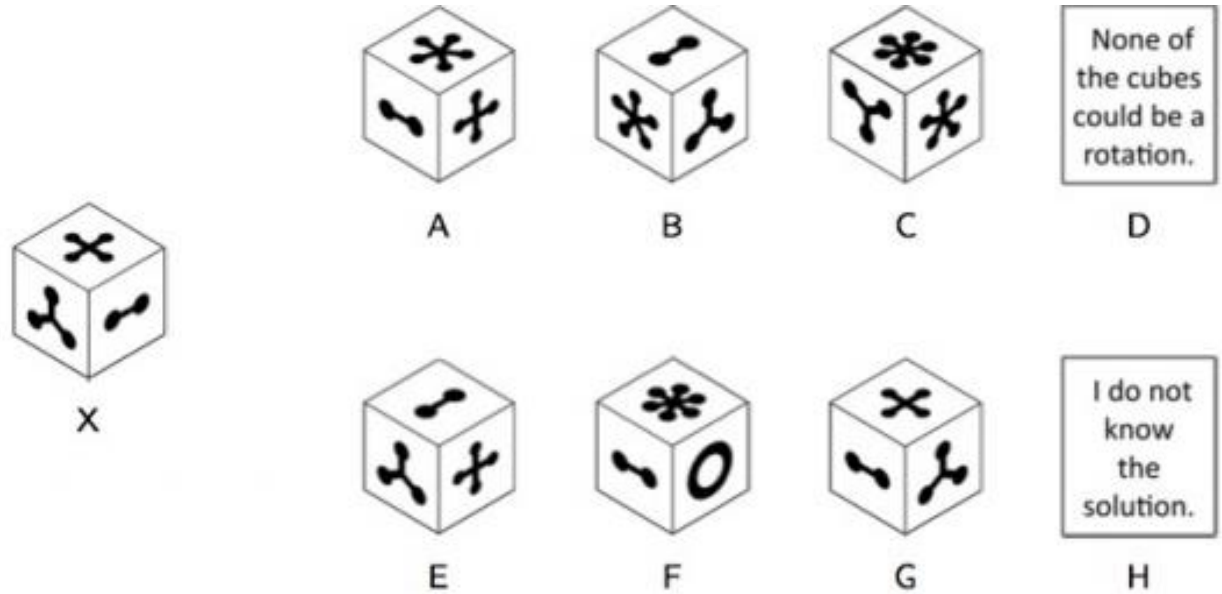
Please indicate which is the best answer to complete the figure below.





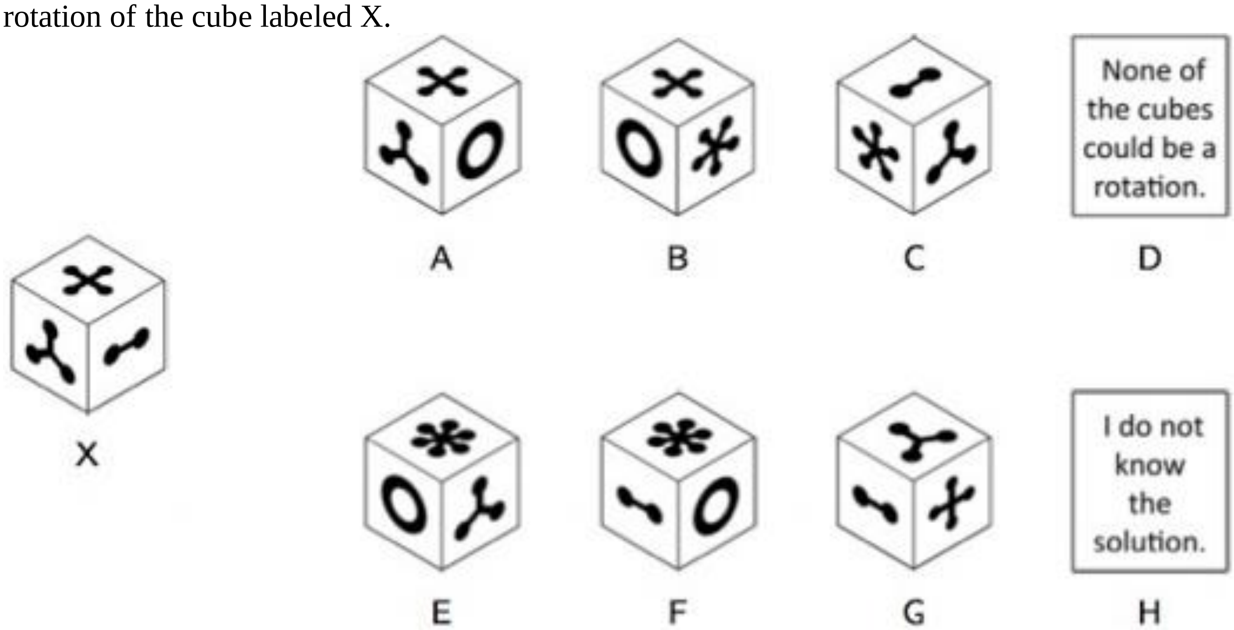
- A
- B
- C
- D
- E
- F

All the cubes below have a different image on each side. Select the choice that could represent a rotation of the cube labeled X.



- B
- C
- D
- E
- F
- G
- H

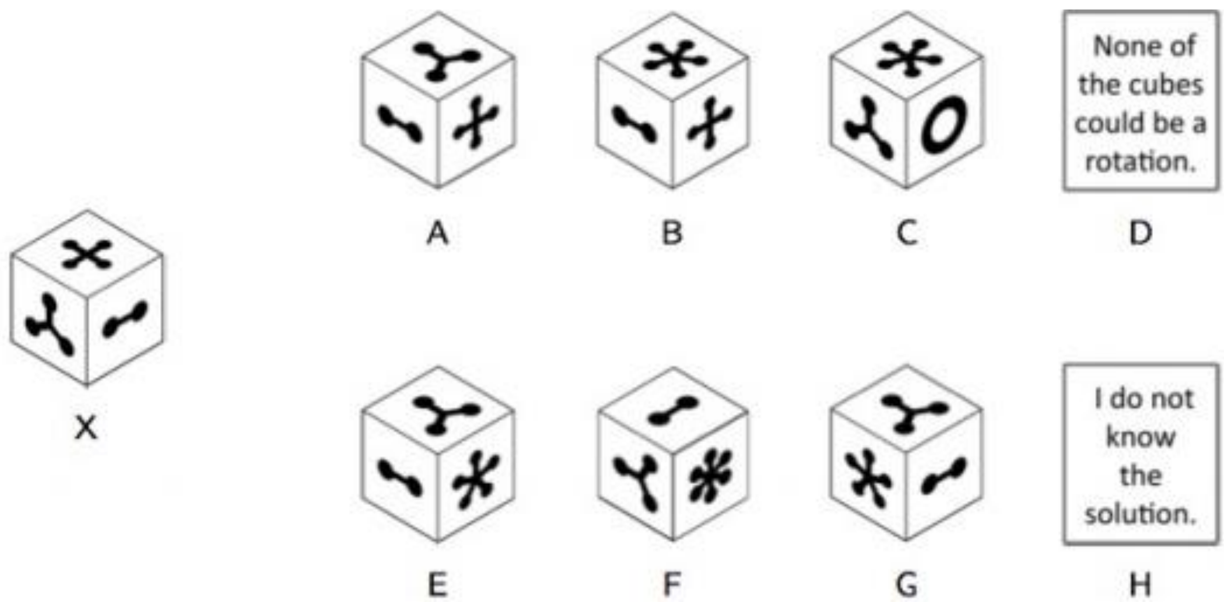
All the cubes below have a different image on each side. Select the choice that could represent a rotation of the cube labeled X.



- A

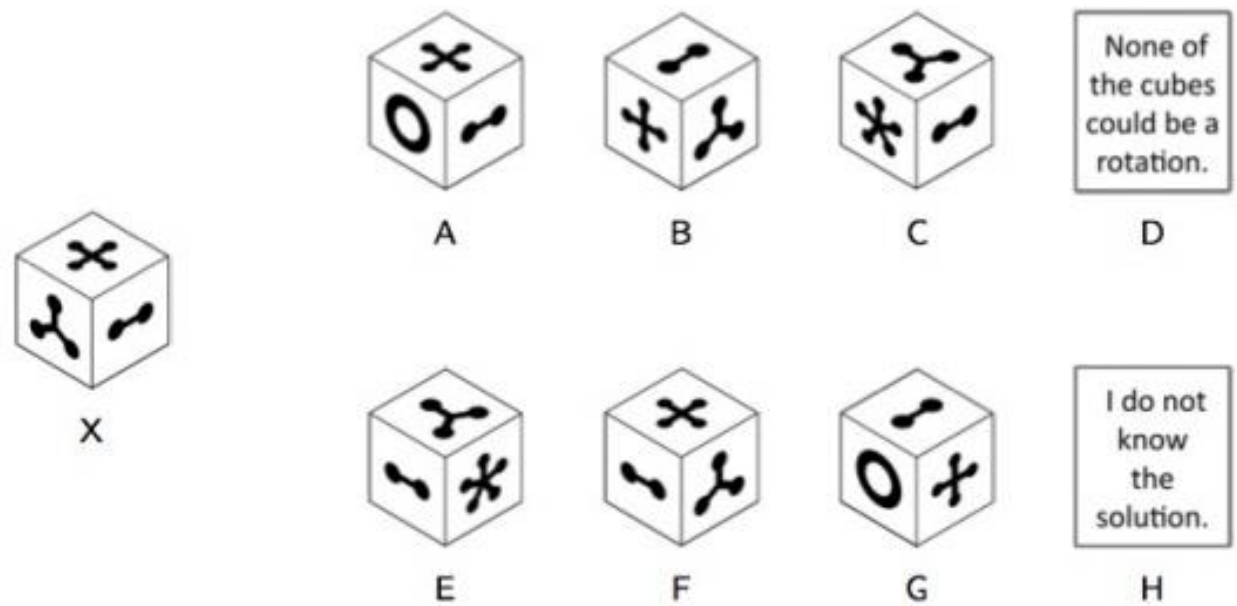
- B
- C
- D
- E
- F
- G
- H

All the cubes below have a different image on each side. Select the choice that could represent a rotation of the cube labeled X.



- A
- B
- C
- D
- E
- F
- G
- H

All the cubes below have a different image on each side. Select the choice that could represent a rotation of the cube labeled X.


A
B
C
None of the cubes could be a rotation.
D
X
E
F
G
I do not know the solution.
H

- A
- B
- C
- D
- E
- F
- G
- H

Appendix F: Research Areas of Application

Cognitive abilities are implicated in every aspect of the human experience. A comprehensive review of research areas in which intelligence plays a role is beyond the scope of this manuscript. Rather, the following section provides a broad overview of a few important areas of application in which an accessible, well-validated measure of cognitive abilities such as the ICAR would be useful to researchers, including academic and professional achievement, and physical and mental health outcomes.

Achievement. General intelligence is among the strongest predictors of both academic (Deary et al., 2007; Floyd et al., 2008; Rohde & Thompson, 2007; Taub et al., 2008) and professional achievement (Kaufman & Kaufman, 2015; Kuncel & Hezlett, 2001; Schmidt & Hunter, 1998). The connection between cognitive ability and achievement in these domains is intuitive; those who have greater cognitive capacity tend to perform better in situations that require learning and adapting to new tasks, like school and work. In fact, several definitions of cognitive ability include the capacity to learn and adapt to novel situations as a core function (Sternberg, 1985; Sternberg & Detterman, 1986; Wechsler, 1975).

Academic Achievement. General intelligence and broad cognitive abilities measured by individual assessments strongly predict achievement scores in math, reading, and writing in children as young as six years of age (Flanagan, Fiorello, & Ortiz, 2010; Floyd et al., 2008; McGrew & Wendling, 2010; Taub et al., 2008). The predictive validity of intelligence constructs persists throughout the education system; scores on tests of cognitive abilities show strong predictive validity for primary and secondary school grades (Laidra, Pullmann, & Allik, 2007; Roth et al., 2015), standardized test scores (Frey & Detterman, 2004) and success in higher education (Barchard, 2003; Kuncel & Hezlett, 2001; Murray & Wren, 2003).

The relation between academic achievement and intelligence remains significant after controlling for relevant factors like socioeconomic status (SES) (Deary et al., 2007; von Stumm, 2017). In a large study of children ages 7 to 16, cognitive ability scores accounted for approximately 40% of the variance in academic performance and growth, while SES only accounted for 2% and 8%, respectively (von Stumm, 2017). Other evidence suggests larger direct effects from SES but the effect of cognitive ability remains a significant predictor after controlling for SES and multiple other relevant covariates (Marks, 2016; Mistry, Biesanz, Chien, Howes, & Benner, 2008; Pearce et al., 2016).

The effect of SES on achievement may be influenced by the development of cognitive abilities in early childhood development (Hackman, Farah, & Meaney, 2010). Lack of access to perinatal care, nutrition, and early cognitive stimulation, hinders cognitive development in early childhood (Appoh & Krekling, 2004; Hackman et al., 2010; Lopez Boo, 2016), to the extent that cognitive ability is almost entirely predicted by environmental factors at low levels of SES (Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). Furthermore, the gap in IQ scores between low and high SES students almost triples from age 7 to age 16 (von Stumm & Plomin, 2015). Despite substantial evidence of the relation between SES and cognitive ability, much remains unknown about the underlying causal mechanisms (Hackman et al., 2010). Measuring and controlling for cognitive abilities in education research will elucidate the complex relationships between factors that influence school learning and cognitive development, and advance the development of programs and policies that advocate for children at all levels of ability.

Professional Achievement. Because cognitive abilities play such an essential role in academic achievement, it is no surprise that both general and specific cognitive abilities appear to affect achievement beyond the education system, in many occupational fields (Bertua, Anderson, & Salgado, 2005; Hunter & Hunter, 1984; Salgado et al., 2003; Schmidt & Hunter, 1998). Cognitive abilities also predict indicators of success related to occupational performance such as income, occupational prestige, and management status (Strenze, 2007; Wai & Rindermann, 2015). The predictive validity of tests of cognitive abilities has led many organizations to incorporate them as part of their personnel selection processes (Cortina, Goldstein, Payne, Davison, & Gilliland, 2000). Despite the success of these measures in talent acquisition, research on cognitive ability measurement in industrial settings has remained relatively stagnant in recent years, and more standardized, theory-driven measures are needed (Scherbaum, Goldstein, Yusko, Ryan, & Hanges, 2012).

The importance of cognitive abilities in academic and professional domains is well-established. However, cognitive ability variables are not always considered in research conducted in these fields. The availability of a psychometrically-sound, theoretically-guided measure of cognitive abilities in the public domain will advance the precision and accuracy of research in fields in which learning is of primary interest, such as work and school.

Health Outcomes. Cognitive epidemiology is emerging as an important line of research that could benefit from accessible cognitive ability measurement tools. A number of studies have demonstrated that general intelligence significantly predicts mortality from a variety of causes (Batty et al., 2008; Der, Batty, & Deary, 2009; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007; Sörberg, Allebeck, & Hemmingsson, 2014). The increased mortality risk associated with low cognitive ability is related to a wide variety of physical and mental health considerations.

Physical Health Outcomes. For example, Batty and colleagues (2008) found cognitive ability predicts mortality from cardiovascular disease at a similar level to smoking, and more strongly than body mass index, total cholesterol, blood pressure or blood glucose levels. This relation continued to be significant even after controlling for confounding factors such as childhood SES. Several other studies have found a significant relation between early intelligence and specific health outcomes later in life, such as unintentional injury (Lawlor, Clark, & Leon, 2007; Whitley et al., 2010), obesity (Yu, Han, Cao, & Guo, 2010), pain (Gale, Deary, Cooper, & Batty, 2012), and addiction (Cadar & Kaushal, 2017).

Mental Health Outcomes. Low cognitive abilities in childhood are also associated with increased risk for mental health problems later in life, including depression, anxiety, autism, and schizophrenia (Gale, Batty, Tynelius, Deary, & Rasmussen, 2010; Zammit et al., 2004). A recent American study found low childhood IQ to be associated with increased relapse and duration of depressive episodes, as well as increase suicidal ideation and attempt later in life (Hung et al., 2015). A longitudinal study of 987,308 Swedish men found that for each unit increase in score on a logic test, the risk of suicide decreased by 12% (Gunnell, Magnusson, & Rasmussen, 2005). The authors suggested the relation between suicidality and low cognitive ability is explained by either by shared etiology of mental health problems and cognitive function, an individual's inability to cope with stressors given limited cognitive resources, or some combination of those causes.

Mechanism of Influence. Although the causal nature of the relation between early life intelligence and mortality remains unclear, several explanations have been speculated. One possibility is that high cognitive abilities are a reflection of overall brain health, or "system integrity" (Deary et al., 2009; Deary, Weiss, & Batty, 2010). This perspective is illustrated well through the lens of aging. Research has demonstrated cognitive decline is a normal part of the aging process, even in the absence of pathology (Salthouse, 2009). Physiological degeneration of the brain leads to cognitive decline, and there is evidence that medical disorders in earlier life increase the risk of early dementia (Brayne, 2007). Conversely, individuals who demonstrate higher levels of cognitive ability early in life, also referred to as cognitive reserve, tend to be able to better defend against both cognitive and physical decline associated with aging and disease (Schwartz, Quaranto, Healy, Benedict, & Vollmer, 2013; Stern, 2009). From this viewpoint, intelligence has a reciprocal relation with overall health and well-being, and is important to measure and track throughout the lifespan.

High cognitive ability may also reduce the likelihood of mortality by promoting health behaviors, as individuals who are more knowledgeable about behavioral risk factors for illness and injury are better equipped to manage their health. This idea is supported by lower rates of engagement in health risk behaviors by individuals with higher cognitive ability. Longitudinal studies demonstrate decreased likelihood of smoking later in life among children who score highly on tests of cognitive abilities (Batty, Deary, & Macintyre, 2007; Martin, Fitzmaurice, Kindlon, & Buka, 2004), and increased likelihood of quitting for those who take up the habit in early adulthood (M. D. Taylor et al., 2003). Adults who binge drink are more likely to have lower cognitive ability scores in childhood than adults who consume alcohol moderately (Batty, Deary, & Macintyre, 2006), and adults with higher concurrent cognitive ability scores are less likely to be overweight or obese (Chandola, Deary, Blane, & Batty, 2006; Deary, Whalley, Batty, & Starr, 2006). In a cohort of 420 children, higher intelligence at age 12 was generally associated with more favorable health behaviors at age 16, such as delay in onset of smoking and less time watching TV (Ciarrochi, Heaven, & Skinner, 2012). Though these relations do not necessarily imply causality, they demonstrate a longitudinal relation between cognitive ability and health behaviors that has ample potential for further research.

Behavioral Interventions. The relation between cognitive abilities and health behaviors yields another important research application of cognitive ability measurement; the influential role of cognitive abilities in behavioral interventions. Health literacy (i.e., the ability to understand health information and make appropriate health-related decisions) is associated with better health outcomes and use of health services (Sheridan et al., 2011). Health literacy has been shown to be highly correlated with early and concurrent cognitive ability (Mõttus et al., 2014). Based on this connection, researchers have suggested that interventions to improve health literacy may be informed by interventions that target general cognitive ability (Nisbett et al., 2012), and prevention and health services may be more effective if tailored to individual cognitive needs (Mõttus et al., 2014). The latter strategy has already been implemented in mental health services, where cognitive behavioral therapy has been successfully adapted for individuals with lower cognitive abilities (Taylor, Lindsay, & Willner, 2008). Though a large amount of research supports the connection between general cognitive ability and physical and mental health outcomes, relatively little has examined its moderating effects on intervention effectiveness (Baker et al., 2007; Mõttus et al., 2014). Barriers to access to psychometrically sound measures that suit the needs of these researchers may result in cognitive abilities being overlooked as an important influence or moderator in the effectiveness of psychoeducational programs.

Summary. Cognitive abilities influence a wide range of fields of research, including education, healthcare, psychology, and industry. The relation between cognitive abilities and individual outcomes in these areas is well established, but much remains unknown about the nature and direction of these relations, and how a better understanding these relations can improve interventions, programs, and public policy. Accessible, validated measurement tools in the public domain will encourage broader study of cognitive abilities as they relate to other areas of individual functioning, and advance research progress across the research community.

Appendix G: A Brief Comment on the Historical Context of Intelligence Theories

The roots of contemporary intelligence theories tend to be traced back to the work of Sir Francis Galton, who used Darwin's ideas to argue that "genius" is a normally distributed and heritable characteristic of humans (Brody, 2000). Galton argued for the existence of racial hierarchies, and is the founder of the field of "eugenics" (Galton, 1883). Though he is thought to be a pioneer in the field, his work is also seen in large part as responsible for the racist and oppressive ideologies underlying the history of intelligence theory and testing in the 19th century, some of which continue on in the discourse around intelligence today (Cravens, 1978; Gould, 1996; Suzuki & Valencia, 1997; Valencia, 2010).

The development of a new test of intelligence would not be complete without acknowledgment of the racially and culturally biased ideologies that dominated the field for most of the 20th century (Valencia, 2010). An early application of intelligence tests in the United States was the evaluation of immigrants at Ellis Island to determine who was fit for various jobs, entry into the country, or to have children (Gould, 1996). Hereditarian views of intelligence have also played a role in the systemic racism of the U.S. education system. Opponents of integration used these arguments to prevent Black students from enrolling in White schools in the mid-1900's (Cook, 1979). These issues persist in the current U.S. education system; disproportionate identification of racial, ethnic, and language minorities in special education (e.g. Artiles & Trent, 1994; Harry, Arnaiz, Klingner, & Sturges, 2008; Morgan et al., 2015; Sullivan & Bal, 2013) has led many scholars to suggest that intelligence testing, and even the concept of intelligence itself, perpetuates systematic bias and institutionalized racism (Blanchett, 2006; Codrington & Fairchild, 2012; Fraser, 2008; Skiba, Poloni-Staudinger, Simmons, Renae Feggins-Azziz, & Chung, 2005).

The full extent to which group-hierarchical ideologies of human intelligence have been applied in oppressive ways to various peoples is extensive, pervasive, and outside the scope of this paper. Still, it is important to recognize the ways in which ideas about human intelligence have been misapplied and misunderstood throughout history in order to avoid these missteps in future work. As history has demonstrated, the study of individual differences has the potential for corruption across the field of psychology when extrapolated to diverse groups of people (Betancourt & López, 1993).

Yet the field need not be abandoned completely to resist cultural bias and oppression. Alternatively, intelligence researchers may recall the original intentions of Alfred Binet, the French psychologist who, along with Simone Theodore, developed the first cognitively based intelligence test (Binet & Simon, 1916). According to Siegler (1992), Binet was compelled to create a more objective assessment in order to help identify children who needed special education services, rather than relying on the (often biased) judgments of teachers and family members. From this perspective, standardized assessments of cognitive abilities have created a fairer system of evaluation for many individuals. But Binet also recognized the limitations of his measure, cautioning users of the test that IQ scores were influenced by a complex interplay of factors variable throughout the lifespan, and thus comparisons of children from different backgrounds are not interpretable (Siegler, 1992). Furthermore, Messick (1979) argues that contemporary intelligence test developers have an ethical responsibility to evaluate not only the validity of assessments in terms of their measurement properties, but also the social consequences of their use.

Though historically practitioners and scholars have not always heeded the cautions of Binet and others, the field of intelligence research has undoubtedly become more rigorous and robust (Brody, 2000; Neisser et al., 1996). Advances in technology have made the evaluation and refinement of intelligence theories and measures more feasible (Cudeck & MacCallum, 2012). As a result, the psychometric approach to defining and understanding the concept of intelligence has prevailed over the last century (Keith & Reynolds, 2012; Naglieri, 2015). Advanced statistical approaches have fostered a number of studies examining racial, ethnic, and gender bias of cognitive assessments (e.g. Keith, 1999; Keith, Quirk, Schartzer, & Elliott, 1999; Keith, Reynolds, Roberts, Winter, & Austin, 2011; Waschl et al., 2016). A better understanding of the systemic racism, sexism, and cultural bias that plagues most American institutions will further improve research quality across the domains of psychology. Despite improvements since the field's inception, it is important to consider the oppressive potential of cognitive testing, especially when considering new applications of cognitive ability measurement such as those mentioned in Appendix F.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30.

Aken, L. van, Kessels, R. P. C., Wingbermühle, E., Veld, W. M. van der, & Egger, J. I. M. (2016). Fluid intelligence and executive functioning more alike than different? *Acta Neuropsychiatrica*, *28*(1), 31–37. https://doi.org/10.1017/neu.2015.46

Akshoomoff, N., Newman, E., Thompson, W. K., McCabe, C., Bloss, C. S., Chang, L., … Jernigan, T. L. (2014). The NIH Toolbox Cognition Battery: Results from a large normative developmental sample (PING). *Neuropsychology*, *28*(1), 1–10. https://doi.org/10.1037/neu0000001

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. *Contemporary Intellectual Assessment: Theories, Tests, And*, (2nd), 185–202.

Alfonso, V. C., Johnson, A., Patinella, L., & Rader, D. E. (1998). Common WISC-III examiner errors: Evidence from graduate students in training. *Psychology in the Schools*, *35*(2), 119–125. https://doi.org/10.1002/(SICI)1520-6807(199804)35:2<119::AID-PITS3>3.0.CO;2-K

Anastasi, A. (1968). *Psychological testing., 3rd ed*. Oxford, England: Macmillan. (1954-15003-000).

Andersen, L. (2014). Visual–Spatial Ability: Important in STEM, Ignored in Gifted Education. *Roeper Review*, *36*(2), 114–121. https://doi.org/10.1080/02783193.2014.884198

Appoh, L. Y., & Krekling, S. (2004). Effects of early childhood malnutrition on cognitive performance of Ghanaian children. *Journal of Psychology in Africa; South of the Sahara,* *the Caribbean, and Afro-Latin America*, *14*(1), 1–7. https://doi.org/10.4314/jpa.v14i1.30604

Artiles, A. J., & Trent, S. C. (1994). Overrepresentation of Minority Students in Special Education: A Continuing Debate. *The Journal of Special Education*, *27*(4), 410–437. https://doi.org/10.1177/002246699402700404

Arvey, R. D. (1972). Some Comments on Culture Fair Tests1. *Personnel Psychology*, *25*(3), 433–448. https://doi.org/10.1111/j.1744-6570.1972.tb00828.x

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *22*(2), 366–377. https://doi.org/10.3758/s13423-014-0699-x

Babcock, R. L., & Laguna, K. (1997). An examination of the factor structure of four of the cognitive abilities included in the educational testing service kit of factor-referenced cognitive tests. *Studies in Educational Evaluation*, *23*(2), 159–168. https://doi.org/10.1016/S0191-491X(97)00010-2

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist*, *56*(11), 851.

Baker, D. W., Wolf, M. S., Feinglass, J., Thompson, J. A., Gazmararian, J. A., & Huang, J. (2007). Health Literacy and Mortality Among Elderly Persons. *Archives of Internal Medicine*, *167*(14), 1503–1509. https://doi.org/10.1001/archinte.167.14.1503

Balboni, G., Naglieri, J. A., & Cubelli, R. (2010). Concurrent and Predictive Validity of the Raven Progressive Matrices and the Naglieri Nonverbal Ability Test. *Journal of Psychoeducational Assessment*, *28*(3), 222–235. https://doi.org/10.1177/0734282909343763

Barchard, K. A. (2003). Does emotional intelligence assist in the prediction of academic success? *Educational and Psychological Measurement*, *63*(5), 840–858.

Bates, T. C., & Gupta, S. (2017). Smart groups of smart people: Evidence for IQ as the origin of collective intelligence in the performance of human groups. *Intelligence*, *60*, 46–56. https://doi.org/10.1016/j.intell.2016.11.004

Batty, G. D., Shipley, M. J., Gale, C. R., Mortensen, L. H., & Deary, I. J. (2008). Does IQ predict total and cardiovascular disease mortality as strongly as other risk factors? Comparison of effect estimates using the Vietnam Experience Study. *Heart*, *94*(12), 1541–1544. https://doi.org/10.1136/hrt.2008.149567

Batty, G. D., Shipley, M. J., Mortensen, L. H., Boyle, S. H., Barefoot, J., Grønbæk, M., … Deary, I. J. (2008). IQ in late adolescence/early adulthood, risk factors in middle age and later all-cause mortality in men: The Vietnam Experience Study. *Journal of Epidemiology & Community Health*, *62*(6), 522–531. https://doi.org/10.1136/jech.2007.064881

Batty, G. David, & Deary, I. J. (2004). Early life intelligence and adult health. *BMJ*, *329*(7466), 585–586. https://doi.org/10.1136/bmj.329.7466.585

Batty, G. David, Deary, I. J., & Gottfredson, L. S. (2007). Premorbid (early life) IQ and Later Mortality Risk: Systematic Review. *Annals of Epidemiology*, *17*(4), 278–288. https://doi.org/10.1016/j.annepidem.2006.07.010

Batty, G. David, Deary, I. J., & Macintyre, S. (2006). Childhood IQ and life course socioeconomic position in relation to alcohol induced hangovers in adulthood: The Aberdeen children of the 1950s study. *Journal of Epidemiology & Community Health*, *60*(10), 872–874. https://doi.org/10.1136/jech.2005.045039

Batty, G. David, Deary, I. J., & Macintyre, S. (2007). Childhood IQ in relation to risk factors for premature mortality in middle-aged persons: The Aberdeen Children of the 1950s study. *Journal of Epidemiology & Community Health*, *61*(3), 241–247. https://doi.org/10.1136/jech.2006.048215

Bauer, P. J., & Zelazo, P. D. (2013). NIH Toolbox Cognition Battery (CB): Summary, Conclusions, and Implications for Cognitive Development. *Monographs of the Society for Research in Child Development*, *78*(4), 133–146. https://doi.org/10.1111/mono.12039

Beaujean, A. A., Hull, D. M., Sheng, Y., Worrell, F. C., Bolen, J., & Verdisco, A. E. (2017). Psychometric properties of the Shipley Block Design Task: A study with Jamaican young adults. *Journal of Psychoeducational Assessment*, *35*(5), 506–520. https://doi.org/10.1177/0734282916643439

Bell, N. L., Rucker, M., Finch, A. j., & Alexander, J. (2002). Concurrent validity of the Slosson full-range intelligence test: Comparison with the Wechsler intelligence scale for children–third edition and the Woodcock Johnson tests of achievement–revised. *Psychology in the Schools*, *39*(1), 31–38. https://doi.org/10.1002/pits.10002

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment*, *22*(1), 121–130. https://doi.org/10.1037/a0017767

Bertua, C., Anderson, N., & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology*, *78*(3), 387–409. https://doi.org/10.1348/096317905X26994

Betancourt, H., & López, S. R. (1993). The study of culture, ethnicity, and race in American psychology. *American Psychologist*, *48*(6), 629.

Bigler, E. D., Johnson, S. C., Jackson, C., & Blatter, D. D. (1995). Aging, brain size and IQ. *Intelligence*, *21*(1), 109–119.

Binet, A., & Simon, T. (1916). *The Development of Intelligence in Children: The Binet-Simon Scale*. Williams & Wilkins Company.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences*, *29*(2), 109–125.

Blanchett, W. J. (2006). Disproportionate representation of african american students in special education: Acknowledging the role of white privilege and racism. *Educational Researcher*, *35*(6), 24–28. https://doi.org/10.3102/0013189X035006024

Borghesani, P. R., Madhyastha, T. M., Aylward, E. H., Reiter, M. A., Swarny, B. R., Warner Schaie, K., & Willis, S. L. (2013). The association between higher order abilities, processing speed, and age are variably mediated by white matter integrity during typical aging. *Neuropsychologia*, *51*(8), 1435–1444. https://doi.org/10.1016/j.neuropsychologia.2013.03.005

Brase, G. L. (2009). How different types of participant payments alter task performance. *Judgment and Decision Making*, *4*(5), 419.

Brayne, C. (2007). The elephant in the room—Healthy brains in later life, epidemiology and public health. *Nature Reviews Neuroscience*, *8*(3), 233–239.

Brody, N. (2000). History of theories and measurements of intelligence. *Handbook of Intelligence*, 16–33.

Cadar, D., & Kaushal, A. (2017). Commentary on Daly & Egan (2017): Intelligence, education and addiction. *Addiction*, *112*(4), 660–661. https://doi.org/10.1111/add.13733

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology Research and Practice*, *31*(2), 141–154.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Campbell, D. T., & O'Connell, E. J. (1982). Methods as diluting trait relationships rather than adding irrelevant systematic variance. *New Directions for Methodology of Social & Behavioral Science*.

Canivez, G. L., & Watkins, M. W. (2010). Investigation of the factor structure of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS–IV): Exploratory and higher order factor analyses. *Psychological Assessment*, *22*(4), 827–836. https://doi.org/10.1037/a0020429

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press.

Cattell, Raymond B. (1971). *Abilities: Their structure, growth, and action*. Oxford, England: Houghton Mifflin. (1973-02450-000).

Cattell, Raymond Bernard. (1987). *Intelligence: Its structure, growth and action* (Vol. 35). Elsevier.

Cattell, Raymond Bernard, & Cattell, A. (1960). *Cattell Culture Fair Intelligence Test: A Measure of" g"*. Bobbs-Merrill.

Chandola, T., Deary, I. J., Blane, D., & Batty, G. D. (2006). Childhood IQ in relation to obesity and weight gain in adult life: The National Child Development (1958) Study.

International Journal of Obesity, *30*(9), 1422–1432. https://doi.org/10.1038/sj.ijo.0803279

Choma, B. L., & Hanoch, Y. (2017). Cognitive ability and authoritarianism: Understanding support for Trump and Clinton. *Personality and Individual Differences*, *106*, 287–291. https://doi.org/10.1016/j.paid.2016.10.054

Ciarrochi, J., Heaven, P. C. L., & Skinner, T. (2012). Cognitive ability and health-related behaviors during adolescence: A prospective study across five years. *Intelligence*, *40*(4), 317–324. https://doi.org/10.1016/j.intell.2012.03.003

Codrington, J., & Fairchild, H. H. (2012). Special education and the mis-education of African American children: A call to action. *The Association of Black Psychologists Washington, DC*.

Condon, D. M., & Revelle, W. (2012, December). *The International Cognitive Ability Resource: Development and initial validation of a public-domain measure*. Presented at the Thirteenth Annual ISIR Conference, San Antonio, TX.

Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*(1), 52–64. https://doi.org/10.1016/j.intell.2014.01.004

Condon, D. M., & Revelle, W. (2016). Selected ICAR data from the SAPA-Project: Development and initial validation of a public-domain measure. *Journal of Open Psychology Data*, *4*(1). https://doi.org/10.5334/jopd.25

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4

Cook, S. W. (1979). Social science and school desegregation: Did we mislead the supreme court? *Personality and Social Psychology Bulletin*, *5*(4), 420–437. https://doi.org/10.1177/014616727900500404

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology*, *53*(2), 325–351. https://doi.org/10.1111/j.1744-6570.2000.tb00204.x

Cravens, H. (1978). The triumph of evolution: American scientists and the heredity-environment controversy, 1900-1941. *22127*. Retrieved from https://repository.library.georgetown.edu/handle/10822/547029

Cronbach, L. J. (1988). Five perspectives on validity argument. *Test Validity*, 3–17.

Cronbach, L. J. (1989). Construct validation after thirty years. *Intelligence: Measurement, Theory, and Public Policy, 3*, 147–171.

Cudeck, R., & MacCallum, R. C. (2012). *Factor Analysis at 100: Historical Developments and Future Directions*. Routledge.

Daniel, M. H. (2012). *Equivalence of Q-interactive Administered Cognitive Tasks: WAIS-IV* [Technical]. Pearson Education, Inc.

Danthiir, V., Roberts, R. D., Pallier, G., & Stankov, L. (2001). What the nose knows: Olfaction and cognitive abilities. *Intelligence*, *29*(4), 337–361.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes: The PASS theory of intelligence*. Needham Heights, MA, US: Allyn & Bacon.

Deary, I. J., Batty, G. D., & Gale, C. R. (2008). Bright children become enlightened adults. *Psychological Science*, *19*(1), 1–6. https://doi.org/10.1111/j.1467-9280.2008.02036.x

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., … Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin*, *92*(1), 135–152. https://doi.org/10.1093/bmb/ldp033

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001

Deary, I. J., Weiss, A., & Batty, G. D. (2010). Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities. *Psychological Science in the Public Interest*, *11*(2), 53–79. https://doi.org/10.1177/1529100610387081

Deary, I. J., Whalley, L. J., Batty, G. D., & Starr, J. M. (2006). Physical fitness and lifetime cognitive change. *Neurology*, *67*(7), 1195–1200. https://doi.org/10.1212/01.wnl.0000238520.06958.6a

Der, G., Batty, G. D., & Deary, I. J. (2009). The association between IQ in adolescence and a range of health outcomes at 40 in the 1979 US National Longitudinal Study of Youth. *Intelligence*, *37*(6), 573–580. https://doi.org/10.1016/j.intell.2008.12.002

Dombrowski, S. C., Canivez, G. L., Watkins, M. W., & Alexander Beaujean, A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition with the 16 primary and secondary subtests. *Intelligence*, *53*, 194–201. https://doi.org/10.1016/j.intell.2015.10.009

Donlon, T. F. (1984). *The College Board technical handbook for the scholastic aptitude test and achievement tests*. College Board.

Dunlop, P. D., Bourdage, J. S., de Vries, R. E., Hilbig, B. E., Zettler, I., & Ludeke, S. G. (2017). Openness to (reporting) experiences that one never had: Overclaiming as an outcome of the knowledge accumulated through a proclivity for cognitive and aesthetic exploration. *Journal of Personality and Social Psychology*, *113*(5), 810–834. https://doi.org/10.1037/pspp0000110

Ekstrom, R. B., & Bejar, I. I. (1990). *Computer-Based Assessment of Cognition: The ETS Factor Kit*. Retrieved from https://eric.ed.gov/?id=ED328587

Elliott, C. D. (2007). Differential Ability Scales—Second Edition. *Harcourt Assessment*, *22*(1), 128–132.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & A, R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Fiske, D. W. (1971). *Measuring the concepts of personality*. Oxford, England: Aldine. (1972-06806-000).

Fiske, D. W., & Campbell, D. T. (1992). Citations do not solve problems. *Psychological Bulletin*, *112*(3), 393–395. https://doi.org/10.1037/0033-2909.112.3.393

Flanagan, D. P., & Alfonso, V. C. (2016). *WJ IV Clinical Use and Interpretation: Scientist-Practitioner Perspectives*. Academic Press.

Flanagan, D. P., Fiorello, C. A., & Ortiz, S. O. (2010). Enhancing practice through application of Cattell–Horn–Carroll theory and research: A “third method” approach to specific learning disability identification. *Psychology in the Schools*, *47*(7), 739–760. https://doi.org/10.1002/pits.20501

Flanagan, D. P., & Harrison, P. L. (2012). *Contemporary Intellectual Assessment, Third Edition: Theories, Tests, and Issues*. Guilford Press.

Floyd, R. G., McGrew, K. S., & Evans, J. J. (2008). The relative contributions of the Cattell-Horn-Carroll cognitive abilities in explaining writing achievement during childhood and adolescence. *Psychology in the Schools*, *45*(2), 132–144. https://doi.org/10.1002/pits.20284

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment*, *7*(3), 248.

Fraser, S. (2008). *The Bell Curve Wars: Race, Intelligence, and the Future of America*. Basic Books.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g?: The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science*, *15*(6), 373–378. https://doi.org/10.1111/j.0956-7976.2004.00687.x

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*(4), 237–241. https://doi.org/10.1111/j.1467-9280.1996.tb00366.x

Fuerst, J., & Kirkegaard, E. O. (2016). Admixture in the Americas: Regional and national differences. *Mankind Quarterly*, *56*(3), 255.

Gale, C. R., Batty, G. D., Tynelius, P., Deary, I. J., & Rasmussen, F. (2010). Intelligence in early adulthood and subsequent hospitalisation and admission rates for the whole range of mental disorders: Longitudinal study of 1,049,663 men. *Epidemiology (Cambridge, Mass.)*, *21*(1), 70–77. https://doi.org/10.1097/EDE.0b013e3181c17da8

Gale, C. R., Deary, I. J., Cooper, C., & Batty, G. D. (2012). Intelligence in childhood and chronic widespread pain in middle age: The National Child Development Survey. *PAIN®*, *153*(12), 2339–2344. https://doi.org/10.1016/j.pain.2012.07.027

Galton, F. (1883). *Inquiries into the human faculty & its development*. JM Dent and Company.

Gambardella, A., & Hall, B. H. (2006). Proprietary versus public domain licensing of software and research products. *Research Policy*, *35*(6), 875–892. https://doi.org/10.1016/j.respol.2006.04.004

Gardner, H. (1987). The theory of multiple intelligences. *Annals of Dyslexia*, *37*(1), 19–35. https://doi.org/10.1007/BF02648057

Garnier-Villarreal, M., Rhemtulla, M., & Little, T. D. (2014). Two-method planned missing designs for longitudinal research. *International Journal of Behavioral Development*, *38*(5), 411–422. https://doi.org/10.1177/0165025414542711

Gerton, B. K., Brown, T. T., Meyer-Lindenberg, A., Kohn, P., Holt, J. L., Olsen, R. K., & Berman, K. F. (2004). Shared and distinct neurophysiological components of the digits forward and backward tasks as revealed by functional neuroimaging. *Neuropsychologia*, *42*(13), 1781–1787. https://doi.org/10.1016/j.neuropsychologia.2004.04.023

Gignac, G. E. (2006). Evaluating subtest ‘g’saturation levels via the single trait-correlated uniqueness (STCU) SEM approach: Evidence in favor of crystallized subtests as the best indicators of ‘g.’ *Intelligence*, *34*(1), 29–46.

Gignac, G. E. (2014). Fluid intelligence shares closer to 60% of its variance with working memory capacity and is a better indicator of general intelligence. *Intelligence*, *47*, 122–133. https://doi.org/10.1016/j.intell.2014.09.004

Gignac, G. E. (2015). Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g. *Intelligence*, *52*, 71–79. https://doi.org/10.1016/j.intell.2015.07.006

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe*, *7*(1), 7–28.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Gong, Q.-Y., Sluming, V., Mayes, A., Keller, S., Barrick, T., Cezayirli, E., & Roberts, N. (2005). Voxel-based morphometry and stereology provide convergent evidence of the importance of medial prefrontal cortex for fluid intelligence in healthy adults. *NeuroImage*, *25*(4), 1175–1186. https://doi.org/10.1016/j.neuroimage.2004.12.044

Gould, S. J. (1996). *The Mismeasure of Man*. W. W. Norton & Company.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323–343. https://doi.org/10.1037/1082-989X.11.4.323

Griffin, P. T., & Heffernan, A. (1983). Digit Span, Forward and Backward: Separate and unequal components of the WAIS Digit Span. *Perceptual and Motor Skills*, *56*(1), 335–338. https://doi.org/10.2466/pms.1983.56.1.335

Gunnell, D., Magnusson, P. K. E., & Rasmussen, F. (2005). Low intelligence test scores in 18 year old men and risk of suicide: Cohort study. *BMJ*, *330*(7484), 167. https://doi.org/10.1136/bmj.38310.473565.8F

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*(3), 179–203. https://doi.org/10.1016/0160-2896(84)90008-4

Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In *Advances in the psychology of human intelligence, Vol. 4* (pp. 35–71). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Gustafsson, J.-E. (2002). Measurement from a hierarchical point of view. *The Role of Constructs in Psychological and Educational Measurement*, 73–95.

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: Mechanistic insights from human and animal research. *Nature Reviews. Neuroscience*, *11*(9), 651–659. https://doi.org/10.1038/nrn2897

Hamby, T., Taylor, W., Snowden, A. K., & Peterson, R. A. (2016). A meta-analysis of the reliability of free and for-pay big five scales. *The Journal of Psychology*, *150*(4), 422–430. https://doi.org/10.1080/00223980.2015.1060186

Harrison, A. G., DeLisle, M. M., & Parker, K. C. (2008). An investigation of the General Abilities Index in a group of diagnostically mixed patients. *Journal of Psychoeducational Assessment*, *26*(3), 247–259.

Harrison, T. L., Shipstead, Z., & Engle, R. W. (2015). Why is working memory capacity related to matrix reasoning tasks? *Memory & Cognition*, *43*(3), 389–396. https://doi.org/10.3758/s13421-014-0473-3

Harry, B., Arnaiz, P., Klingner, J., & Sturges, K. (2008). Schooling and the construction of identity among minority students in spain and the united states. *The Journal of Special Education*, *42*(1), 15–25. https://doi.org/10.1177/0022466907313605

Heitz, R. P., Redick, T. S., Hambrick, D. Z., Kane, M. J., Conway, A. R. A., & Engle, R. W. (2006). Working memory, executive function, and general fluid intelligence are not the same. *Behavioral and Brain Sciences*, *29*(2), 135–136. https://doi.org/10.1017/S0140525X06319036

Hessl, D., Sansone, S. M., Berry-Kravis, E., Riley, K., Widaman, K. F., Abbeduto, L., … Gershon, R. C. (2016). The NIH Toolbox Cognitive Battery for intellectual disabilities: Three preliminary studies and future directions. *Journal of Neurodevelopmental Disorders*, *8*(1), 35. https://doi.org/10.1186/s11689-016-9167-4

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. *Woodcock-Johnson Technical Manual*, 197–232.

Horn, J. L., & Blankson, N. (2005). *Foundations for Better Understanding of Cognitive Abilities.*

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253.

Hossiep, R., Turck, D., & Hasella, M. (1999). *Bochumer Matrizentest (BOMAT) Advanced*. Hogrefe.

Houghton Mifflin Harcourt. (2017). Retrieved July 11, 2017, from http://www.hmhco.com/hmh-assessments/clinical-and-special-needs-assessment/wj-iv/shop-now?&atrkid=V3ADW6700B927_29430318052_kwd-59284400642__189715403680_g_c___1t1&gclid=EAIaIQobChMIzOjPubCB1QIVgSSBCh1u0g4cEAAYASAAEgIuRvD_BwE

Hung, G. C.-L., Pietras, S. A., Carliner, H., Martin, L., Seidman, L. J., Buka, S. L., & Gilman, S. E. (2015). Cognitive ability in childhood and the chronicity and suicidality of depression. *The British Journal of Psychiatry*, bjp.bp.114.158782. https://doi.org/10.1192/bjp.bp.114.158782

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98. https://doi.org/10.1037/0033-2909.96.1.72

Jia, F., Moore, E. W. G., Kinai, R., Crowe, K. S., Schoemann, A. M., & Little, T. D. (2014). Planned missing data designs with small sample sizes: How small is too small? *International Journal of Behavioral Development*, *38*(5), 435–452. https://doi.org/10.1177/0165025414531095

Johnsen, S. K. (2017). *Test of nonverbal intelligence: A language-free measure of cognitive ability*. 185–206. https://doi.org/10.1007/978-3-319-50604-3_11

Johnson, W., & Bouchard Jr., T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*(4), 393–416. https://doi.org/10.1016/j.intell.2004.12.002

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *32*(1), 95–107. https://doi.org/10.1016/S0160-2896(03)00062-X

Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, *36*(1), 81–95.

Kail, R. (2000). Speed of information processing. *Journal of School Psychology*, *38*(1), 51–61. https://doi.org/10.1016/S0022-4405(99)00036-9

Kail, R., & Hall, L. K. (2001). Distinguishing short-term memory from working memory. *Memory & Cognition*, *29*(1), 1–9. https://doi.org/10.3758/BF03195735

Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D. M., Gajda, A., … Benedek, M. (2016). Is creativity without intelligence possible? A Necessary Condition Analysis. *Intelligence*, *57*, 105–117. https://doi.org/10.1016/j.intell.2016.04.006

Kaufman, A. S. (2004). *KABC-II: Kaufman Assessment Battery for Children*. AGS Pub.

Kaufman, A. S., & Kaufman, N. L. (1990). *K-BIT: Kaufman Brief Intelligence Test: Manual*. American Guidance Service.

Kaufman Assessment Battery for Children, Second Edition. (2017). Retrieved July 11, 2017, from Western Psychological Services website: https://www.wpspublish.com/store/p/2828/kabc-ii-kaufman-assessment-battery-for-children-second-edition

Kaufman, J. C., & Kaufman, A. S. (2001). Time for the changing of the guard: A farewell to short forms of intelligence tests. *Journal of Psychoeducational Assessment*, *19*(3), 245–267. https://doi.org/10.1177/073428290101900305

Kaufman, J. C., & Kaufman, A. S. (2015). It can be very tempting to throw out the baby with the bathwater: A father-and-son commentary on “does iq really predict job performance?” *Applied Developmental Science*, *19*(3), 176–181. https://doi.org/10.1080/10888691.2015.1008922

Kaufmann, P. M. (2009). Protecting raw data and psychological tests from wrongful disclosure: A primer on the law and other persuasive strategies. *The Clinical Neuropsychologist*, *23*(7), 1130–1159. https://doi.org/10.1080/13854040903107809

Kaya, F., Delen, E., & Bulut, O. (2012). Test Review: Shipley-2 Manual. *Journal of Psychoeducational Assessment*, *30*(6), 593–597. https://doi.org/10.1177/0734282912440852

Keith, T. Z. (1999). Effects of general and specific abilities on student achievement: Similarities and differences across ethnic groups. *School Psychology Quarterly*, *14*(3), 239.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-: What does it measure? *School Psychology Review*, *35*(1), 108.

Keith, T. Z., Low, J. A., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2010). Higher-order factor structure of the Differential Ability Scales–II: Consistency across ages 4 to 17. *Psychology in the Schools*, *47*(7), 676–697. https://doi.org/10.1002/pits.20498

Keith, T. Z., Quirk, K. J., Schartzer, C., & Elliott, C. D. (1999). Construct bias in the Differential Ability Scales? Confirmatory and hierarchical factor structure across three ethnic groups. *Journal of Psychoeducational Assessment*, *17*(3), 249–268.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we’ve learned from 20 years of research. *Psychology in the Schools*, *47*(7), 635–650. https://doi.org/10.1002/pits.20496

Keith, T. Z., & Reynolds, M. R. (2012). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, P. L. Harrison, D. P. Flanagan (Ed), & P. L. Harrison (Ed) (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues.* (pp. 758–799). New York, NY, US: Guilford Press. (2012-09043-032).

Keith, T. Z., Reynolds, M. R., Roberts, L. G., Winter, A. L., & Austin, C. A. (2011). Sex differences in latent cognitive abilities ages 5 to 17: Evidence from the Differential Ability Scales—Second Edition. *Intelligence*, *39*(5), 389–404.

Krach, S. K., Loe, S. A., Jones, W. P., & Farrally, A. (2009). Convergent validity of the reynolds intellectual assessment scales (rias) using the woodcock—Johnson tests of cognitive ability, third edition (wj-iii) with university students. *Journal of Psychoeducational Assessment*, *27*(5), 355–365. https://doi.org/10.1177/0734282909331749

Kuncel, N. R., & Hezlett, S. A. (2001). *Academic performance, career potential, creativity, and job performance: Can one construct predict them all?* https://doi.org/10.1037/0022-3514.86.1.148

Kvist, A., & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence*, *36*(5), 422–436. https://doi.org/10.1016/j.intell.2007.08.004

Kyllonen, P. C., & Dennis, A. (1996). Is working memory capacity Spearman's g. *Human Abilities: Their Nature and Measurement*, 49–75.

Lager, E., Melin, B., Hemmingsson, T., & Sörberg Wallin, A. (2017). The evolving relationship between premorbid intelligence and serious depression across the lifespan—A longitudinal study of 43,540 Swedish men. *Journal of Affective Disorders*, *211*, 37–43. https://doi.org/10.1016/j.jad.2016.12.051

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences*, *42*(3), 441–451. https://doi.org/10.1016/j.paid.2006.08.001

Lawlor, D. A., Clark, H., & Leon, D. A. (2007). Associations between childhood intelligence and hospital admissions for unintentional injuries in adulthood: The aberdeen children of the 1950s cohort study. *American Journal of Public Health*, *97*(2), 291–297. https://doi.org/10.2105/AJPH.2005.080168

Leffard, S. A., Miller, J. A., Bernstein, J., DeMann, J. J., Mangis, H. A., & McCoy, E. L. (2006). Substantive validity of working memory measures in major cognitive functioning test batteries for children. *Applied Neuropsychology*, *13*(4), 230–241.

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines*, *17*(4), 391–444. https://doi.org/10.1007/s11023-007-9079-x

Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV Assessment*. John Wiley & Sons.

Lichtenberger, E. O., & Kaufman, A. S. (2012). *Essentials of WAIS-IV Assessment, Second Edition* (2nd ed., Vol. 96). John Wiley & Sons.

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*(4), 199–204. https://doi.org/10.1111/cdep.12043

Lodge, J. (2012). The Concurrent Validity of the Shipley-2 and the WAIS-IV. *Browse All Theses and Dissertations*. Retrieved from http://corescholar.libraries.wright.edu/etd_all/652

Lohman, D. F. (1979). *Spatial Ability: A Review and Reanalysis of the Correlational Literature.* Stanford University of California School of Education.

Lohman, D. F. (1996). Spatial ability and g. *Human Abilities: Their Nature and Measurement*, *97*, 116.

Lohman, D. F., Thorndike, R. L., Hagen, E. P., Smith, P., Fernandes, C., & Strand, S. (2001). Cognitive abilities test. *Windsor, England: NFER-Nelson*.

Lopez Boo, F. (2016). Socio-economic status and early childhood cognitive skills: A mediation analysis using the Young Lives panel. *International Journal of Behavioral Development*, *40*(6), 500–508. https://doi.org/10.1177/0165025416644689

Lowe, N. K., & Ryan-Wenger, N. M. (1992). Beyond Campbell and Fiske: Assessment of convergent and discriminant validity. *Research in Nursing & Health*, *15*(1), 67–75.

Marks, G. N. (2016). The relative effects of socio-economic, demographic, non-cognitive and cognitive influences on student achievement in Australia. *Learning and Individual Differences*, *49*, 1–10. https://doi.org/10.1016/j.lindif.2016.05.012

Martin, L. T., Fitzmaurice, G. M., Kindlon, D. J., & Buka, S. L. (2004). Cognitive performance in childhood and early adult illness: A prospective cohort study. *Journal of Epidemiology & Community Health*, *58*(8), 674–679. https://doi.org/10.1136/jech.2003.016444

Martínez, K., & Colom, R. (2009). Working memory capacity and processing efficiency predict fluid but not crystallized and spatial intelligence: Evidence supporting the neural noise hypothesis. *Personality and Individual Differences*, *46*(3), 281–286. https://doi.org/10.1016/j.paid.2008.10.012

McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, Present, and Future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp. 136–181). New York, NY, US: Guilford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

McGrew, K. S., & Evans, J. J. (2004). Internal and external factorial extensions to the Cattell–Horn–Carroll (CHC) theory of cognitive abilities: A review of factor analytic research since Carroll's Seminal 1993 Treatise. *Institute for Applied Psychometrics*.

McGrew, K. S., Flanagan, D. P., Keith, T. Z., & Vanderwood, M. (1997). Beyond g: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review*.

McGrew, K. S., & Wendling, B. J. (2010). Cattell–Horn–Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, *47*(7), 651–675. https://doi.org/10.1002/pits.20497

Messick, S. (1979). Test Validity and the Ethics of Assessment. *ETS Research Report Series*, *1979*(1), i–43. https://doi.org/10.1002/j.2333-8504.1979.tb01178.x

Mistry, R. S., Biesanz, J. C., Chien, N., Howes, C., & Benner, A. D. (2008). Socioeconomic status, parental investments, and the cognitive and behavioral outcomes of low-income children from immigrant and native households. *Early Childhood Research Quarterly*, *23*(2), 193–212. https://doi.org/10.1016/j.ecresq.2008.01.002

Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions. *Educational Researcher*, *44*(5), 278–292. https://doi.org/10.3102/0013189X15591157

Motta, R. W., & Joseph, J. M. (2000). Group Intelligence Tests. In *Handbook of Psychological Assessment* (pp. 131–146).

Mõttus, R., Johnson, W., Murray, C., Wolf, M. S., Starr, J. M., & Deary, I. J. (2014). Towards understanding the links between health literacy and physical health. *Health Psychology*, *33*(2), 164–173. https://doi.org/10.1037/a0031439

Mrazik, M., Janzen, T. M., Dombrowski, S. C., Barford, S. W., & Krawchuk, L. L. (2012). Administration and scoring errors of graduate students learning the wisc-iv: Issues and controversies. *Canadian Journal of School Psychology*, *27*(4), 279–290. https://doi.org/10.1177/0829573512454106

Mungas, D., Widaman, K., Zelazo, P. D., Tulsky, D., Heaton, R. K., Slotkin, J., … Gershon, R. C. (2013). NIH Toolbox Cognition Battery (CB): Factor Structure for 3 to 15 Year Olds. *Monographs of the Society for Research in Child Development*, *78*(4), 103–118. https://doi.org/10.1111/mono.12037

Murray, C., & Wren, C. T. (2003). Cognitive, academic, and attitudinal predictors of the grade point averages of college students with learning disabilities. *Journal of Learning Disabilities*, *36*(5), 407–415. https://doi.org/10.1177/00222194030360050201

Naglieri, J. A. (2015). Hundred years of intelligence testing: Moving from traditional IQ to second-generation intelligence tests. In S. Goldstein, D. Princiotta, J. A. Naglieri, S. Goldstein (Ed), D. Princiotta (Ed), & J. A. Naglieri (Ed) (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts.* (pp. 295–316). https://doi.org/10.1007/978-1-4939-1562-0_20

National Science Foundation. (2017). Retrieved from www.nsf.com

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., … others. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77.

NIH Toolbox. (n.d.). Retrieved September 15, 2018, from http://www.healthmeasures.net/explore-measurement-systems/nih-toolbox

Niileksela, C. R., Reynolds, M. R., & Kaufman, A. S. (2013). An alternative Cattell–Horn–Carroll (CHC) factor structure of the WAIS-IV: Age invariance of an alternative model for ages 70–90. *Psychological Assessment*, *25*(2), 391–404. https://doi.org/10.1037/a0031175

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130–159. https://doi.org/10.1037/a0026699

Nussbeck, F. W., Eid, Michael., & Lischetzke, Tanja. (2006). Analysing multitrait–multimethod data with structural equation models for ordinal variables applying the WLSMV estimator: What sample size is needed for valid results? *British Journal of Mathematical and Statistical Psychology*, *59*(1), 195–213. https://doi.org/10.1348/000711005X67490

Oakland, T., Glutting, J., & Watkins, M. W. (2005). Assessment of test behaviors with the WISC-IV. *WISC-IV Clinical Use and Interpretation: Scientist Practitioner Perspectives*, 435–463.

Olchowski, A. E. (2008). *Assessing the impact of physical conditioning, dietary intake, body fat, and tobacco use on blood pressure parameters: A two-method measurement design approach*. ProQuest Information & Learning, US. (2008-99080-179).

Pearce, A., Sawyer, A. C. P., Chittleborough, C. R., Mittinty, M. N., Law, C., & Lynch, J. W. (2016). Do early life cognitive ability and self-regulation skills explain socio-economic inequalities in academic achievement? An effect decomposition analysis in UK and Australian cohorts. *Social Science & Medicine*, *165*, 108–118. https://doi.org/10.1016/j.socscimed.2016.07.016

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. (2005). The General (g), Broad, and Narrow CHC Stratum Characteristics of the WJ III and WISC-III Tests: A Confirmatory Cross-Battery Investigation. *School Psychology Quarterly*, *20*(1), 66.

Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Raven, J. C. (1998). *Raven's progressive matrices*. Oxford Psychologists Press Oxford.

Revelle, W., Condon, D. M., Doebler, P., Holling, H., Rust, J., Stillwell, D., … Loe, A. (2014). The International Cognitive Ability Resource Team. Retrieved July 7, 2017, from The International Cognitive Abilities Resource Team website: https://icar-project.com/

Reynolds, C. R. (1997). Forward and backward memory span should not be combined for clinical analysis. *Archives of Clinical Neuropsychology*, *12*(1), 29–40. https://doi.org/10.1093/arclin/12.1.29

Reynolds, M. R., Hajovsky, D. B., Pace, J. R., & Niileksela, C. R. (2016). What does the shipley-2 measure for children and adolescents? Integrated and conjoint confirmatory factor analysis with the wisc-iv. *Assessment*, *23*(1), 23–41. https://doi.org/10.1177/1073191115572695

Reynolds, M. R., & Keith, T. Z. (2007). Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence*, *35*(3), 267–281. https://doi.org/10.1016/j.intell.2006.08.002

Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology*, *51*(4), 535–555. https://doi.org/10.1016/j.jsp.2013.02.003

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Robinson, D. L. (1999). The IQ factor: Implications for intelligence theory and measurement. *Personality and Individual Differences*, *27*(4), 715–735.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, *35*(1), 83–92. https://doi.org/10.1016/j.intell.2006.05.004

Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition*, *52*(3), 326–333. https://doi.org/10.1016/S0278-2626(03)00170-2

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. https://doi.org/10.1016/j.intell.2015.09.002

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the european community. *Journal of Applied Psychology*, *88*(6), 1068–1081. https://doi.org/10.1037/0021-9010.88.6.1068

Salthouse, T. A. (2009). Decomposing age correlations on neuropsychological and cognitive variables. *Journal of the International Neuropsychological Society: JINS*, *15*(5), 650–661. https://doi.org/10.1017/S1355617709990385

Salthouse, T. A. (2014). Evaluating the correspondence of different cognitive batteries. *Assessment*, *21*(2), 131–142. https://doi.org/10.1177/1073191113486690

Sanders, S., McIntosh, D. E., Dunham, M., Rothlisberg, B. A., & Finch, H. (2007). Joint confirmatory factor analysis of the differential ability scales and the Woodcock-Johnson Tests of Cognitive Abilities–Third Edition. *Psychology in the Schools*, *44*(2), 119–138. https://doi.org/10.1002/pits.20211

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations*. JM Sattler San Diego, CA.

Scherbaum, C. A., Goldstein, H. W., Yusko, K. P., Ryan, R., & Hanges, P. J. (2012). Intelligence 2.0: Reestablishing a research program on g in I–O psychology. *Industrial and Organizational Psychology*, *5*(2), 128–148.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173. https://doi.org/10.1037/0022-3514.86.1.162

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262

Schmitt, N., & Stults, D. M. (1986). Methodology Review: Analysis of Multitrait-Multimethod Matrices. *Applied Psychological Measurement*, *10*(1), 1–22. https://doi.org/10.1177/014662168601000101

Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review*, *25*(1), 12–27. https://doi.org/10.1016/j.hrmr.2014.09.004

Schneider, W., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. *Contemporary Intellectual Assessment: Theories, Tests, And*, (3rd), 99–144.

Schrank, F. A., Decker, S. L., & Garruto, J. M. (2016). *Essentials of WJ IV Cognitive Abilities Assessment*. John Wiley & Sons.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). Woodcock-Johnson IV Tests of Achievement. *Rolling Meadows, IL: Riverside*.

Schwartz, C. E., Quaranto, B. R., Healy, B. C., Benedict, R. H., & Vollmer, T. L. (2013). Cognitive Reserve and Symptom Experience in Multiple Sclerosis: A Buffer to Disability Progression Over Time? *Archives of Physical Medicine and Rehabilitation*, *94*(10), 1971-1981.e1. https://doi.org/10.1016/j.apmr.2013.05.009

Sheridan, S. L., Halpern, D. J., Viera, A. J., Berkman, N. D., Donahue, K. E., & Crotty, K. (2011). Interventions for Individuals with Low Health Literacy: A Systematic Review. *Journal of Health Communication*, *16*(sup3), 30–54. https://doi.org/10.1080/10810730.2011.604391

Shipley, W. C., Gruber, C. P., & Martin, K. (2009). *Shipley-2: Manual*. Los Angeles, CA: Western Psychology Services.

Siegler, R. S. (1992). The other Alfred Binet. *Developmental Psychology*, *28*(2), 179.

Skiba, R. J., Poloni-Staudinger, L., Simmons, A. B., Renae Feggins-Azziz, L., & Chung, C.-G. (2005). Unproven links: Can poverty explain ethnic disproportionality in special education? *The Journal of Special Education*, *39*(3), 130–144. https://doi.org/10.1177/00224669050390030101

Sörberg, A., Allebeck, P., & Hemmingsson, T. (2014). IQ and somatic health in late adolescence. *Intelligence*, *44*, 155–162. https://doi.org/10.1016/j.intell.2014.04.002

Spearman, C., & Jones, L. W. (1950). *Human ability*. London: Macmillan.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015–28. https://doi.org/10.1016/j.neuropsychologia.2009.03.004

Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. CUP Archive.

Sternberg, R. J. (1992). Psychological Bulletin's top 10 "hit parade." *Psychological Bulletin*, *112*(3), 387–388. https://doi.org/10.1037/0033-2909.112.3.387

Sternberg, R. J., & Detterman, D. K. (1986). *What is intelligence?: Contemporary viewpoints on its nature and definition*. Praeger Pub Text.

Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist*, *38*(8), 878–893. https://doi.org/10.1037/0003-066X.38.8.878

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. Oxford University Press.

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*(5), 401–426. https://doi.org/10.1016/j.intell.2006.09.004

Styck, K. M., & Walsh, S. M. (2016). Evaluating the prevalence and impact of examiner errors on the Wechsler scales of intelligence: A meta-analysis. *Psychological Assessment*, *28*(1), 3–17. https://doi.org/10.1037/pas0000157

Sullivan, A. L., & Bal, A. (2013). Disproportionality in Special Education: Effects of Individual and School Variables on Disability Risk. *Exceptional Children*, *79*(4), 475–494. https://doi.org/10.1177/001440291307900406

Suzuki, L. A., & Valencia, R. R. (1997). Race–ethnicity and measured intelligence: Educational implications. *American Psychologist*, *52*(10), 1103–1114. https://doi.org/10.1037/0003-066X.52.10.1103

Taub, G. E., Keith, T. Z., Floyd, R. G., & Mcgrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly*, *23*(2), 187–198. https://doi.org/10.1037/1045-3830.23.2.187

Taylor, J. L., Lindsay, W. R., & Willner, P. (2008). Cbt for people with intellectual disabilities: Emerging evidence, cognitive ability and IQ effects. *Behavioural and Cognitive Psychotherapy*. Retrieved from /core/journals/behavioural-and-cognitive-psychotherapy/article/cbt-for-people-with-intellectual-disabilities-emerging-evidence-cognitive-ability-and-iq-effects/A9BA555977843E816DD07EF959123E89

Taylor, M. D., Hart, C. L., Smith, G. D., Starr, J. M., Hole, D. J., Whalley, L. J., … Deary, I. J. (2003). Childhood mental ability and smoking cessation in adulthood: Prospective observational study linking the Scottish Mental Survey 1932 and the Midspan studies. *Journal of Epidemiology & Community Health*, *57*(6), 464–465. https://doi.org/10.1136/jech.57.6.464

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*(6), 623–628. https://doi.org/10.1046/j.0956-7976.2003.psci_1475.x

Undheim, J. O., & Gustafsson, J.-E. (1987). The hierarchical organization of cognitive abilities: Restoring general intelligence through the use of linear structural relations (LISREL). *Multivariate Behavioral Research*, *22*(2), 149–171.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038

Valencia, R. R. (2010). *Dismantling Contemporary Deficit Thinking: Educational Thought and Practice*. Routledge.

Vanderwood, M. L., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (2002). The contribution of general and specific cognitive abilities to reading achievement. *Learning and Individual Differences*, *13*(2), 159–188.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence*, *7*(1), 53–70. https://doi.org/10.1016/0160-2896(83)90006-5

von Stumm, S. (2017). Socioeconomic status amplifies the achievement gap throughout compulsory education independent of intelligence. *Intelligence*, *60*, 57–62. https://doi.org/10.1016/j.intell.2016.11.006

von Stumm, S., & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence, 48,* 30–36. https://doi.org/10.1016/j.intell.2014.10.002

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817.

Wai, J., & Rindermann, H. (2015). The path and performance of a company leader: A historical examination of the education and cognitive ability of Fortune 500 CEOs. *Intelligence*, *53*, 102–107. https://doi.org/10.1016/j.intell.2015.10.001

Ward, L. C., Bergman, M. A., & Hebert, K. R. (2012). WAIS-IV subtest covariance structure: Conceptual and statistical considerations. *Psychological Assessment*, *24*(2), 328–340. https://doi.org/10.1037/a0025614

Waschl, N. A., Nettelbeck, T., Jackson, S. A., & Burns, N. R. (2016). Dimensionality of the Raven's Advanced Progressive Matrices: Sex differences and visuospatial ability. *Personality and Individual Differences*, *100*, 157–166. https://doi.org/10.1016/j.paid.2015.12.008

Wechsler, D. (1999). Manual for the Wechsler abbreviated intelligence scale (WASI). *San Antonio, TX: The Psychological Corporation*.

Wechsler, David. (1975). Intelligence defined and undefined: A relativistic appraisal. *American Psychologist*, *30*(2), 135–139. https://doi.org/10.1037/h0076868

Wechsler, David. (2008). *Wechsler Adult Intelligence Scale–fourth edition technical and interpretive manual.* San Antonio, TX: Pearson.

Wechsler, David. (2014). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV)*.

Wechsler Intelligence Scale for Children®-Fifth Edition. (2017). Retrieved July 11, 2017, from Pearson Clinical website: http://www.pearsonclinical.com/psychology/products/100000771/wechsler-intelligence-scale-for-childrensupsupfifth-edition--wisc-v.html/?utm_source=google&utm_medium=ppc&utm_campaign=3009632-D-701b00000006KgP&cmpid=701b00000006KgP

Weintraub, S., Dikmen, S., K Heaton, R., Tulsky, D., Zelazo, P., Slotkin, J., … Gershon, R. (2014). The Cognition Battery of the NIH Toolbox for Assessment of Neurological and Behavioral Function: Validation in an Adult Sample. *Journal of the International Neuropsychological Society : JINS*, *20*, 1–12. https://doi.org/10.1017/S1355617714000320

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WAIS-IV and Clinical Validation of the Four- and Five-Factor Interpretative Approaches. *Journal of Psychoeducational Assessment*, *31*(2), 94–113. https://doi.org/10.1177/0734282913478030

West, T. G. (1991). *In the mind's eye: Visual thinkers, gifted people with learning difficulties, computer images, and the ironies of creativity.* Prometheus Books.

Whitley, E., Batty, G. D., Gale, C. R., Deary, I. J., Tynelius, P., & Rasmussen, F. (2010). Intelligence in early adulthood and subsequent risk of unintentional injury over two decades: Cohort study of 1 109 475 Swedish men. *Journal of Epidemiology & Community Health*, *64*(5), 419–425. https://doi.org/10.1136/jech.2009.100669

Wilhelm, O. (2005). Measuring reasoning ability. *Handbook of Understanding and Measuring Intelligence*, 373–392.

Woodcock, R. W. (1990). Theoretical Foundations of the Wj-R Measures of Cognitive Ability. *Journal of Psychoeducational Assessment*, *8*(3), 231–258. https://doi.org/10.1177/073428299000800303

Wothke, W. (1996). Models for multitrait-multimethod matrix analysis. *Advanced Structural Equation Modeling: Issues and Techniques*, 7–56.

Yates, B. T., & Taub, J. (2003). Assessing the costs, benefits, cost-effectiveness, and cost-benefit of psychological assessment: We should, we can, and here's how. *Psychological Assessment*, *15*(4), 478–495. https://doi.org/10.1037/1040-3590.15.4.478

Yu, Z. B., Han, S. P., Cao, X. G., & Guo, X. R. (2010). Intelligence in relation to obesity: A systematic review and meta-analysis. *Obesity Reviews*, *11*(9), 656–670. https://doi.org/10.1111/j.1467-789X.2009.00656.x

Zammit, S., Allebeck, P., David, A. S., Dalman, C., Hemmingsson, T., Lundberg, I., & Lewis, G. (2004). A Longitudinal Study of Premorbid IQ Score and Risk of Developing Schizophrenia,Bipolar Disorder, Severe Depression, and Other Nonaffective Psychoses. *Archives of General Psychiatry*, *61*(4), 354–360. https://doi.org/10.1001/archpsyc.61.4.354